

LEVEL REDUCTION AND THE QUANTUM THRESHOLD THEOREM

Thesis by
Panos Aliferis

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy



California Institute of Technology
Pasadena, California

2007

(Defended December 11, 2006)

arXiv:quant-ph/0703230v1 25 Mar 2007

Copyright notice:

Chapters 2 and 3 contain some material from [1] and chapter 4 contains some material from [2] which are both copyrighted by Rinton Press. Chapter 5 contains some material from [3] which is copyrighted by the American Physical Society.

The remaining material is

©2007

Panos Aliferis

All Rights Reserved

for Ayis and K., of course

Foreword

Living on the open air of the mountains of research and reaching the point of writing this thesis would not have been possible without the guidance, the encouragement, and the support of many people. To them I write the few thanking words that follow.

First, I would like to thank my advisor John Preskill. He provided the vessel in the form of IQI and the guidance for my education at Caltech. From him I learned much. Most importantly, I learned how the successful scientific spirit combines depth with breadth, seriousness with cheerfulness, abstract proofs with happy dancing. He introduced me to the field of fault-tolerant quantum computation and pointed out interesting problems in this area of research. Many of the results of this thesis came about from my collaboration with him and Daniel Gottesman and would not have been possible without their crucial contributions and insights.

Second, I would like to thank my mentor and friend Debbie Leung. I was fortunate to meet her as a starting graduate student at Caltech and collaborate with her on my first research results. From her I also learned much. I remember her emphasizing in several different occasions the value of developing easily composable frameworks for attacking research problems, and it was inspired by these words that I was guided to the idea of level reduction. Her sincere and active interest about both my personal well-being and education were invaluable in keeping me balanced at many difficult times.

Third, I would like to thank my collaborators David DiVincenzo and Barbara Terhal for sharing many thoughts and starting new projects with me. My visits to the IBM Quantum Information group in the summers of 2005 and 2006 gave me new ideas and helped me understand in more depth important elements of our common research.

Fourth, I would like to thank my collaborators Andrew Cross and Krysta Svore with whom I exchanged many ideas on various topics of this thesis. Andrew carried out the threshold calculations discussed in chapter 5 and welcomed my ideas and suggestions with an enthusiasm that is so characteristic of him. Krysta contributed in the software we used for these calculations and, with her fresh smile and positive attitude, was a source of encouragement and inspiration.

Fifth, I would like to thank my fellow graduate students, Kovid Goyal, Hui Khoon Ng, Graeme Smith, and Mike Zwolak, who spent time discussing their research with me. I also owe thanks to

Robin Blume-Kohout, Sergei Bravyi, David Poulin, and Robert Raussendorf whom I met as postdocs at IQI and with whom I have had many insightful discussions.

Finally, I do not forget the influence of my three pre-Caltech mentors: my math mentor in Sparta, Stavros Damopoulos, my advisor at NTUA, Heraklis Avramopoulos, and my advisor at the University of Michigan, Duncan Steel. I also appreciate the help of the members of my defense and candidacy committees, Alexei Kitaev, Hideo Mabuchi, Robert McEliece, and Chris Umans.

Pasadena, XII 06.

Abstract

Computers have led society to the information age revolutionizing central aspects of our lives from production and communication to education and entertainment. There exist, however, important problems which are intractable with the computers available today and, experience teaches us, will remain so even with the more advanced computers we can envision for tomorrow.

Quantum computers promise speedups to some of these important but classically intractable problems. Simulating physical systems, a problem of interest in a diverse range of areas from testing physical theories to understanding chemical reactions, and solving number factoring, a problem at the basis of cryptographic protocols that are used widely today on the internet, are examples of applications for which quantum computers, when built, will offer a great advantage over what is possible with classical computer technology.

The construction of a quantum computer of sufficient scale to solve interesting problems is, however, especially challenging. The reason for this is that, by its very nature, operating a quantum computer will require the coherent control of the quantum state of a very large number of particles. Fortunately, the theory of quantum error correction and fault-tolerant quantum computation gives us confidence that such quantum states can be created, can be stored in memory and can also be manipulated provided the quantum computer can be isolated to a sufficient degree from sources of noise.

One of the central results in the theory of fault-tolerant quantum computation, the quantum threshold theorem shows that a noisy quantum computer can accurately and efficiently simulate any ideal quantum computation provided that noise is weakly correlated and its strength is below a critical value known as the quantum *accuracy threshold*. This thesis provides a simpler and more transparent non-inductive proof of this theorem based on the concept of *level reduction*. This concept is also used in proving the quantum threshold theorem for coherent and leakage noise and for quantum computation by measurements. In addition, the proof provides a methodology which allows us to establish improved rigorous lower bounds on the value of the quantum accuracy threshold.

Contents

Foreword	iv
Abstract	vi
1 Prologue	1
1.1 The Quantum Computer	1
1.2 Introduction to Quantum Computation	2
1.2.1 Quantum Mechanics and Qubits	3
1.2.2 Open Systems and the Theory of Physical Operations	6
1.2.3 The Quantum Circuit Model	8
1.2.4 Accuracy	10
1.2.5 Quantum Universality	14
1.3 Quantum Coding	15
1.3.1 Quantum Error-Correction Criteria	17
1.3.2 Stabilizer Codes	19
1.3.3 CSS Construction	21
1.3.4 Examples	22
2 The Syntax of Fault-Tolerant Quantum Computation	24
2.1 Introduction	24
2.2 Local Noise	25
2.3 Fault-Tolerant Quantum Circuit Simulations	28
2.4 Accuracy	34
2.5 History and Acknowledgements	39
3 Recursive Simulations and the Quantum Threshold Theorem	40
3.1 Introduction	40
3.2 Recursive Fault-Tolerant Simulations	41
3.3 The Level-Reduction Lemma	43

3.3.1	Independent Bad Rectangles as Simulated Faults	45
3.3.2	The Combinatorics for Overlapping Rectangles	48
3.4	The Quantum Threshold Theorem for Local Noise	48
3.4.1	Improving the Threshold Estimate	50
3.5	History and Acknowledgements	52
4	Further Applications of Level Reduction	54
4.1	Introduction	54
4.2	Leakage Noise	55
4.2.1	The Local Leakage Noise Model	56
4.2.2	The Leakage-Reduction Lemma	57
4.2.2.1	Bad Level-0 Rectangles As Simulated Faults	60
4.2.3	The Quantum Threshold Theorem for Local Leakage Noise	61
4.2.4	Contracting Level-0 Rectangles	62
4.3	Measurement-Based Quantum Computation	63
4.3.1	Gate Teleportation and Quantum Computation by Measurements	64
4.3.2	The Model-Conversion Lemma	66
4.3.3	The Threshold Theorem for Quantum Computation by Measurements	68
4.4	History and Acknowledgements	69
5	Lower Bounds on the Quantum Accuracy Threshold	70
5.1	Introduction	70
5.2	Gadget Constructions	71
5.2.1	Fault-Tolerant Quantum Error Correction	71
5.2.1.1	Syndrome Measurement with Cat States	71
5.2.1.2	Syndrome Measurement with Encoded Pauli Eigenstates	74
5.2.1.3	Syndrome Measurement via Logical Teleportation	75
5.2.1.4	Satisfying Property 1(a)	76
5.2.2	Encoded Quantum Universality	78
5.2.2.1	Encoded Clifford-Group Operations	78
5.2.2.2	Quantum Software and Encoded Non-Clifford Gates	79
5.2.3	Fault-Tolerant Preparation and Measurement	80
5.3	Quantum Fault Tolerance with the Bacon-Shor Code	80
5.3.1	Syndrome Measurement using the Gauge Qubits	82
5.3.2	Standard Syndrome Measurement Procedures	84
5.3.3	Accuracy Threshold Lower Bounds	85
5.3.3.1	The 9-Qubit Bacon-Shor Code	86

5.3.3.2	The 25-Qubit Bacon-Shor Code	89
5.3.3.3	Summary of Results	91
5.3.4	Universal Quantum Computation	92
5.3.4.1	The Logical Phase Gate	92
5.3.4.2	The Logical Toffoli Gate	95
5.3.4.3	Bounds on the Accuracy of Quantum Software	100
5.4	History and Acknowledgements	103
6	Epilogue	104
A	Deriving the Bacon-Shor code	105
A.1	Starting from Shor's Code	105
A.2	Starting from Bravyi and Kitaev's Surface Code	105
B	Teleporting into Code Blocks	107
	Bibliography	109

Chapter 1

Prologue

1.1 The Quantum Computer

Trying to introduce the concept of quantum computation, it is tempting to begin by saying that quantum computers are computing devices that operate according to the laws of quantum mechanics. But, on second thought, this description is not very informative since quantum mechanics, in so far as it is a complete natural theory, describes every physical computing device and, so, even classical computers. What distinguishes quantum from classical computers is that the operation of the former is based on the two distinctively quantum-mechanical effects of *interference* and *entanglement* that do not appear in classical physics. So this is how I would describe in one phrase quantum computers to those who have taken a basic course in quantum mechanics: a quantum computer is an interference device of many entangled computation paths. Just as an interference pattern can appear by preparing a particle in a superposition of different geometric paths which are then combined to interfere, the output of a quantum computer is obtained by preparing the quantum bits in a superposition of different classical computation states which are also combined to interfere producing the final computation answer.

This immediately leads to the problem of dealing with noise in quantum computation. If we think of a quantum computer as a big interference device, it is important to ask whether it is practically possible to construct such a device in a way that the arbitrarily complex interference effect will not disappear due to the fact that the components of the computer are not perfectly accurate. After all, we know that the interference pattern in a double-slit experiment disappears if it is *in principle* possible to know from which slit the particle has passed. So, if we think of a quantum computer as executing a computation which resembles a very complex interference experiment, it is natural to worry that the complexity of the computation will make it impossible *in practice* to maintain the complex superpositions of states and observe an interference effect at the end of the computation.

But before I address the problem of constructing robust quantum computers which is the main topic of this thesis, I should first mention the indications we have about their greater computational

power relative to classical computers. Historically, it was first Feynman [4, 5] and Deutsch [6, 7] who realized that simulating certain quantum systems on classical computers apparently requires time that scales exponentially with the size of the simulated system (e.g, the number of particles), and suggested that a more natural simulator for quantum systems is a computer which is quantum itself. Mathematical models of quantum computation were also previously discussed by Manin [8], by Benioff [9], and by Peres [10]. The computational power of quantum computers has been investigated extensively since these first proposals. Today, proposed quantum computing applications range from general-purpose quantum simulations and simulators of local quantum field theories (e.g., [11, 12, 13, 14, 15]), to understanding the physics of condensed-matter systems and computing ground-state energies of complex molecules (e.g., [16, 17, 18]). But certainly, the strongest evidence for the power of quantum computers comes from Shor’s discovery of a polynomial-time quantum algorithm [19] for finding the prime factors of composite numbers and computing the discrete logarithm. Although there is no proof, it is believed that no polynomial-time classical algorithm exists for these two problems. In fact, the belief in the hardness of these problems has led to their use in the public-key cryptography systems RSA and Diffie-Hellman which are widely used on the internet. More recent results on quantum algorithms include Grover’s quantum algorithm [20] for unstructured search problems which offers quadratic speedups over classical search algorithms and Hallgren’s quantum algorithm [21] for Pell’s equation which is exponentially faster than any known classical algorithm.

The remaining of this introductory chapter is structured as follows: In §1.2, I give a short introduction to quantum computation. I begin by discussing the formalism of quantum mechanics and the theory of physical operations. Then, I introduce the circuit model of quantum computation, including the standard notation for elementary quantum operations and quantum circuit diagrams. Next, I discuss how imperfections in implementing the elementary operations during the execution of a quantum algorithm relate to the accuracy of the final computation result. Finally, I introduce the notion of universal quantum computation and discuss how a quantum computer being able to execute gates from a finite gate set can efficiently simulate any quantum computation to any desired accuracy. In §1.3, I briefly present the main results in the theory of quantum error correction. I begin by discussing the differences between classical and quantum coding and then I state the formal criteria for quantum error correction. Finally, I introduce the stabilizer formalism that enables a succinct description of a large class of quantum codes and use this formalism to describe CSS quantum codes which are especially advantageous for fault-tolerant quantum computation.

1.2 Introduction to Quantum Computation

Computers are physical devices that run algorithms. Intuitively, an algorithm is a set of instructions that given some input produce as an output the answer to some problem. In fact, the *Church-*

Turing thesis—really, a hypothesis about the nature of computers—states that for any function that we intuitively consider computable there exists a corresponding algorithm computing the function that runs on a (universal) Turing machine given sufficient time and memory space (see, e.g., Sipser’s introductory book on the theory of computation [22]). The interesting content of this thesis is that it identifies our intuitive notion of what it means to be able to compute a function with the algorithms that Turing machines can run. As a corollary to this thesis, any computer is as computationally powerful as a universal Turing machine if we do not consider any time or memory limitations.

Even though any computer can theoretically run any algorithm, efficiency considerations are important for determining which algorithms can be executed in practice. To classify algorithms according to their efficiency, one typically considers how the time and memory space for running an algorithm depend on the size of the input to the algorithm, where the size of the input is measured in the *natural units* for the problem considered (e.g., number of binary digits when the input is an integer). Another fundamental thesis in the theory of computation, the *strong Church-Turing thesis*, states that (probabilistic) Turing machines can efficiently simulate any other realistic model of computation. In other words, any realistic computing device can simulate any other one with a slowdown which scales at most as a polynomial with the size of the input. The word *realistic* should be understood as referring to computing devices that can be constructed and they are scalable, i.e., can be extended to receive inputs of any size; e.g., one should probably exclude analog classical computers whose power relies on manipulating real numbers of unlimited precision and which are believed to be extremely vulnerable to inaccuracies and noise.

Quantum computers pose a challenge to the strong Church-Turing thesis. In particular, as I very briefly discussed in the previous section, there are problems for which efficient quantum algorithms exist and no efficient classical algorithms are known. Certainly, at this point one might ask whether quantum computation classifies as a realistic model of computation. To address this question, I will next first describe the standard quantum circuit model of computation. Then, I will discuss the issue of accuracy and universality in quantum computation. As we shall see, the power of quantum computers stems from the parallelism that characterizes the quantum world, and this power is not lost if we consider the limited precision with which we can execute the elementary operations in a quantum algorithm.

1.2.1 Quantum Mechanics and Qubits

Nobody who is not acquainted with the quantum theory and the basic notions of quantum computation should expect to learn them from my thesis. Nevertheless, it would be inappropriate to start without briefly reviewing the basic formalism of quantum mechanics and some basic results from the theory of physical operations—I advise the uninitiated reader to study the first chapters in Nielsen and Chuang’s book [23] and, going into more depth, Preskill’s lectures notes [24] and Kitaev, Shen

and Vyalyi's book [25].

Quantum mechanics is a mathematical theory for the description (and prediction) of outcomes of experiments on physical systems. Let us first consider *closed* physical systems, i.e., systems which can be described in complete isolation from their surroundings. Quantum mechanics associates with such a closed system a *quantum state*, $|\psi\rangle$, which is a *vector* of unit length in a—finite-dimensional, for our purposes—complex inner-product vector space \mathcal{H} (i.e., a Hilbert space). Overall phases are irrelevant, i.e., $\forall \theta \in \mathbb{R}, \forall |\psi\rangle \in \mathcal{H}, e^{i\theta}|\psi\rangle$ describes the same state as $|\psi\rangle$.

Consider the space of bounded operators $\mathcal{B}(\mathcal{H})$ defined in \mathcal{H} . To each measured quantity, x , there corresponds a self-adjoint (i.e., Hermitian) operator or *observable*, $\hat{x} \in \mathcal{B}(\mathcal{H})$, such that the possible outcomes when a measurement of this quantity is performed are given by the eigenvalues, $\{x_i\}$, of \hat{x} . The probability of obtaining the i th outcome, x_i , when a measurement of x is performed on a system prepared in the state $|\psi\rangle$ is given by

$$\text{Prob}(x_i) = \|P_{x_i}|\psi\rangle\|^2 = \langle\psi|P_{x_i}|\psi\rangle, \quad (1.1)$$

where $\langle\psi| \equiv (|\psi\rangle)^\dagger$ († denotes Hermitian conjugation), and

$$P_{x_i} \equiv \sum_{j: \hat{x}|x_j\rangle = x_i|x_j\rangle} |x_j\rangle\langle x_j| \quad (1.2)$$

is the projector onto the eigenspace of \hat{x} corresponding to eigenvalue x_i . After the measurement, the system is described by the state $P_{x_i}|\psi\rangle$ up to normalization—we say that the measurement *collapses* the initial state $|\psi\rangle$ to the state $P_{x_i}|\psi\rangle$. The requirement that \hat{x} is Hermitian implies that all its eigenvalues are real and its eigenvectors form an orthonormal basis spanning \mathcal{H} .

Given the state $|\psi(t_o)\rangle$ at some time t_o , its time evolution is governed by the Schrödinger equation

$$\partial_t |\psi(t)\rangle = -iH|\psi(t)\rangle, \quad (1.3)$$

where ∂_t denotes the partial derivative with respect to time, H is a Hermitian operator (the *Hamiltonian* of the system) and $\hbar = 1$. Equation (1.3) implies that the system dynamics is unitary, i.e., the quantum state at time t is $|\psi(t)\rangle = U|\psi(t_o)\rangle$, where $U = \exp(-iHt)$ is a unitary operator.

In analogy with classical computation where the basic carriers of information are bits, in quantum computation the carriers of information are quantum bits or simply *qubits*. Qubits are two-dimensional quantum systems such as, e.g., spin 1/2 particles.¹ A basis for operators in a two-

¹Depending on the actual experimental situation, a qubit can be formed by any two levels in a multilevel quantum system; e.g., a qubit can be defined by the ground and first excited state of an ion.

dimensional Hilbert space can be formed by the three *Pauli matrices*,

$$X \equiv \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} ; \quad Y \equiv \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} ; \quad Z \equiv \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} , \quad (1.4)$$

and the identity operator $I \equiv \sigma_0 = \text{diag}(1, 1)$. A useful observation is that the Pauli matrices are unitary and Hermitian. The Pauli matrices are, in fact, related to the operators $J_i = \frac{1}{2}\sigma_i$ which generate the rotation group $SU(2)$ in two dimensions: a rotation around the unit-vector $\vec{n} \in \mathbb{R}^3$ by an angle $\theta \in \mathbb{R}$ is given by $U_{\vec{n}}(\theta) = \exp(-i \vec{n} \cdot \vec{J} \theta) = \cos(\theta/2)I - i \vec{n} \cdot \vec{\sigma} \sin(\theta/2)$, where $\vec{\sigma} \equiv (\sigma_1, \sigma_2, \sigma_3)$.

Unlike the state of a bit which is either 0 or 1, the state of a qubit is a unit vector in a two-dimensional Hilbert space, $\mathcal{H} = \mathbb{C}^2$. According to the standard convention, we can choose a basis for \mathbb{C}^2 to be formed by the eigenvectors of the Pauli Z operator: one basis state is the $+1$ eigenvector which we label $|0\rangle$ and the other basis state is the -1 eigenvector which we label $|1\rangle$. Then, an arbitrary single-qubit state can be written in the form $|\psi\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$, where $\alpha_0, \alpha_1 \in \mathbb{C}$ and $|\alpha_0|^2 + |\alpha_1|^2 = 1$ in order for $|\psi\rangle$ to be normalized to have unit length. The basis $\{|0\rangle, |1\rangle\}$ (and its obvious generalization to more than one qubits) we will often refer to as the *computation basis* since it corresponds to the basis along which classical computation is phrased.

Qubits generalize classical bits since not only can they be in the two orthogonal states $|0\rangle$ and $|1\rangle$, but they can also be in *superpositions* of these two states. Similarly, if we consider some number $n > 1$ of qubits, the quantum state, $|\psi^{(n)}\rangle$, that describes them is a unit vector in a 2^n -dimensional Hilbert space; hence, it can in general be written as a superposition of all n -bit strings,

$$|\psi^{(n)}\rangle = \sum_{i_0=0}^1 \sum_{i_1=0}^1 \cdots \sum_{i_{n-1}=0}^1 \alpha_{i_0, \dots, i_{n-1}} |i_0 i_1 \cdots i_{n-1}\rangle , \quad (1.5)$$

where $\alpha_{i_0, \dots, i_{n-1}} \in \mathbb{C}$ are arbitrary *amplitude* coefficients subject to the constraint that $|\psi^{(n)}\rangle$ is normalized. Quantum states such as $|\psi^{(n)}\rangle$ allow us to talk about *quantum parallelism* since the n qubits appear to be in all 2^n possible classical bit configurations *in parallel*. And since both the norms and phases of the amplitude coefficients are important to determine a quantum state, we often say that $|\psi^{(n)}\rangle$ is a *coherent* superposition of the different computation-basis states (as opposed to a probabilistic superposition that could result by picking each computation-basis state at random according to some probability distribution).

In equation (1.5) we have chosen to express $|\psi^{(n)}\rangle$ in terms of the n -qubit computation basis, $\{|i_0 i_1 \cdots i_{n-1}\rangle\}$. In general, a different orthonormal basis could have been used which would have resulted in different amplitude coefficients. In fact, it could be the case that for a particular choice of basis all amplitudes become zero except for a single one. Then, $|\psi^{(n)}\rangle$ expressed in this basis is just a tensor product of single-qubit quantum states, i.e., $|\psi^{(n)}\rangle = |\psi_0\rangle \otimes |\psi_1\rangle \otimes \cdots \otimes |\psi_{n-1}\rangle$

for some states $\{|\psi_i\rangle\}$; such states are called *product states*. Quantum states which are not product states—i.e., states for which, independent of the choice of orthonormal basis, at least two amplitudes are non-zero—are called *entangled*.

1.2.2 Open Systems and the Theory of Physical Operations

Although as illustrated by the Gottesman-Knill theorem [42, 23] entanglement is not sufficient by itself to guarantee that a quantum computation cannot be efficiently simulated on a classical computer, generating and manipulating highly entangled states is certainly *necessary* to giving quantum computers their greater power². Unfortunately, it is a fact that entanglement is very fragile and noise or systematic hardware imperfections tend to induce *decoherence* by destroying quantum superpositions.

To obtain a general description of decoherence processes, it is convenient to consider the qubits of our quantum computer as comprising an *open* physical system with Hilbert space \mathcal{H}_S which is part of a larger closed system with Hilbert space \mathcal{H}_{SB} including all sources of noise. (Often, we will call \mathcal{H}_S the *system* and its “complement” \mathcal{H}_B inside \mathcal{H}_{SB} the *environment* or *bath*; then, $\mathcal{H}_{SB} = \mathcal{H}_S \otimes \mathcal{H}_B$.) The quantum state describing the open system is, in general, no longer a unit vector in \mathcal{H}_S . Instead, it is an operator, $\rho \in \mathcal{B}(\mathcal{H}_S)$, which we call a *density matrix*. Density matrices are Hermitian, positive semidefinite and have unit trace. Just as in closed systems, the evolution of the density matrix of an open system due to dynamics that only involve the open system itself is unitary: when the Hamiltonian of the open system is H_S , the unitary evolution for time t is given by $U_S = \exp(-iH_S t)$ and the initial density matrix ρ evolves to $U_S \rho U_S^\dagger$. Also, the probability of obtaining the i th outcome, x_i , when the quantity x is measured is $\text{Prob}(x_i) = \text{Tr}(P_{x_i} \rho)$, and the measurement leaves the open system with the post-measurement density matrix $P_{x_i} \rho P_{x_i}^\dagger$ up to normalization.

The density-matrix formalism generalizes the formalism describing closed systems since the latter can be recovered if one identifies $\rho = |\psi\rangle\langle\psi|$. Moreover, the density-matrix formalism has the advantage of allowing us to describe parts of a closed system without giving a description of the dynamics in the entire closed system. In particular, restricting our attention to \mathcal{H}_S , it is possible to obtain a useful characterization of the most general evolution of its state which is consistent with some “reasonable” requirements about such an evolution. This characterization is given by the Kraus representation theorem that is stated next.

Theorem 1 (Kraus Representation Theorem [26, 24]). *Let ρ be the initial density matrix of a system with Hilbert space \mathcal{H}_S . Let a set of operators $\{M_k\}$ with $M_k \in \mathcal{B}(\mathcal{H}_S)$ such that $\sum_k M_k^\dagger M_k = I_S$*

²Although, of course, there is no proof but only strong evidence that BQP, i.e., bounded-error polynomial-time quantum computation, is strictly more powerful than BPP, i.e., bounded-error probabilistic polynomial-time classical computation.

(where I_S is the identity operator in \mathcal{H}_S). Then, $\{M_k\}$ defines the quantum operation, or super-operator,

$$\mathcal{E} : \rho \rightarrow \sum_k M_k \rho M_k^\dagger. \quad (1.6)$$

Furthermore, for any linear quantum operation, \mathcal{E} , that satisfies the requirements that (i) $\mathcal{E}(\rho)$ is Hermitian, (ii) $\mathcal{E}(\rho)$ has unit trace, and (iii) $\mathcal{E}(\rho \otimes I_C)$ is positive semidefinite (where I_C is the identity operator in any other Hilbert space, \mathcal{H}_C , different than \mathcal{H}_S), there exists a set, $\{M_k\}$, of Kraus operators as defined above such that the operation can be written in the form (1.6).

Requirements (i) and (ii) are consequences of requiring that $\mathcal{E}(\rho)$ be a density matrix. Requirement (iii) is known as the requirement for *complete positivity*. If we would forget about \mathcal{H}_C , then (iii) would just be the requirement that $\mathcal{E}(\rho)$ is positive semidefinite as is required for any density matrix. Complete positivity imposes the additional, innocuous looking, requirement that \mathcal{E} evolves the density matrix $\rho \otimes I_C$ for any \mathcal{H}_C decoupled from \mathcal{H}_S also to a density matrix.

Superoperators and Stochastic Processes

Imagine a quantum operation, \mathcal{N} , describing noise acting on our quantum computer which, at some point during the computation, is in the state ρ . We often consider a noise model where one of the Kraus operators is proportional to the identity operator I_S , $M_0 = \sqrt{1-p} I_S$ for some real constant $0 \leq p \leq 1$. All other nonidentity Kraus operators will satisfy $\sum_{k \neq 0} M_k^\dagger M_k = p I_S$. In this situation, we can interpret \mathcal{N} as a *stochastic* process which with probability p applies some error to our quantum state ρ and with probability $1-p$ leaves the state unchanged.

To make the correspondence between \mathcal{N} and a stochastic process more concrete, let us consider an additional system, the “bath,” with Hilbert space \mathcal{H}_B such that the dimension of \mathcal{H}_B is at least equal to the number of Kraus operators describing \mathcal{N} . Letting $\{|k\rangle_B\}$ be an orthonormal basis in \mathcal{H}_B , we can consider the isometry defined by

$$U_{SB} : \rho \otimes |0\rangle\langle 0|_B \rightarrow \sum_{k,l} M_k \rho M_l^\dagger \otimes |k\rangle\langle l|_B, \quad \forall \rho \in \mathcal{B}(\mathcal{H}_S). \quad (1.7)$$

U_{SB} can be extended to a unitary in $\mathcal{H}_S \otimes \mathcal{H}_B$ and, furthermore, if we trace over the bath we find

$$\text{Tr}_B \left(U_{SB} (\rho \otimes |0\rangle\langle 0|_B) U_{SB}^\dagger \right) = \sum_k M_k \rho M_k^\dagger, \quad (1.8)$$

where $M_k \equiv {}_B\langle k|U_{SB}|0\rangle_B$. We can therefore interpret \mathcal{N} as describing a process in which our system interacts with some bath in an initial state $|0\rangle_B$ via the unitary U_{SB} , followed by a measurement on the bath along the orthonormal basis $\{|k\rangle_B\}$ whose outcome is *unknown*; outcome k occurs with probability $M_k^\dagger M_k$, in which case the post-measurement system state is proportional to $M_k \rho M_k^\dagger$.

In particular, if $M_0 = \sqrt{1-p} I_S$, then the outcome $k = 0$ occurs with probability $1 - p$ and the post-measurement state in \mathcal{H}_S is the initial state ρ , consistent with our interpretation in the previous paragraph.

1.2.3 The Quantum Circuit Model

After this brief introduction to quantum mechanics and the problem of decoherence, it is time to describe the standard circuit model of quantum computation. Since quantum circuits generalize classical circuits, it is natural to start by describing classical computation in terms of circuits.

A classical circuit is a representation of a Boolean function as a composition of other, more elementary Boolean functions. We consider boolean functions—also often called *gates*—of the form $f : \mathbb{Z}_2^n \rightarrow \mathbb{Z}_2$; i.e., they take a n -bit input to a one-bit output. Let us first fix some set, \mathcal{G} , of gates which we will call our *gate set* or *basis*. Gates in \mathcal{G} play the role of the abstract elementary hardware components of a general-purpose classical computer, and they each have a fixed constant number of input bits. Then, a circuit for computing the Boolean function $g : \mathbb{Z}_2^n \rightarrow \mathbb{Z}_2$ is a finite sequence g_1, g_2, \dots, g_m of gates in \mathcal{G} such that $g(y) = (g_m \circ \dots \circ g_2 \circ g_1)(y)$, $\forall y \in \mathbb{Z}_2^n$. In order to be well defined, this gate sequence should be represented as an acyclic directed graph, or *circuit*. The vertices of the graph correspond to the gates and edges denote bits; the inputs to a gate are a subset of the bits of the input y , a subset of the output bits from preceding gates, and possibly some ancillary bits; the outputs of a gate are either input to succeeding gates or are part of the final computation output. A gate set, \mathcal{G} , is called *universal* if for every Boolean function g , there exists a circuit using gates from \mathcal{G} that computes g . Examples of universal gate sets are the sets $\{\neg, \wedge\}$, $\{\neg, \vee\}$, and $\{\Lambda^2(X)\}$; here $\forall a, b, c \in \mathbb{Z}_2$, $\neg(a) = a + 1$, $\wedge(a, b) = a \cdot b$, $\vee(a, b) = \neg(\wedge(\neg a, \neg b))$, $\Lambda^2(X)(a, b, c) = (a, b, (a \cdot b) + c)$, and arithmetic is done modulo 2.

To go from classical circuits to quantum circuits we first need to change from *bits* to *qubits*:³ a quantum computer processes the state of some finite number, n , of qubits which are each initialized in some fixed single-qubit state, say $|0\rangle$. Quantum computation proceeds by applying a unitary transformation to these qubits, followed by a final measurement along the computation basis on some subset of them. In analogy to classical circuits, a quantum circuit is a representation of a unitary transformation that acts on n qubits as a finite sequence of elementary unitary transformations or, simply, *quantum gates* which operate on a fixed constant number of qubits. These elementary quantum gates play the role of the basic hardware operations of our quantum computer and they can be chosen from a *finite* universal gate set which we may again denote by \mathcal{G} .

In a straightforward generalization of this definition of quantum circuits, we may allow initializing each qubit in a different single-qubit state. We may also allow intermediate measurements whose

³The discussion in this and the following chapters could be generalized if d -dimensional systems, *qudits*, were used instead. But, for simplicity, we will restrict ourselves to discussing qubits.

outcome is used to control quantum gates on other qubits, and also measurements along different bases than the computation basis. However, we do *not* allow initialization in more complicated multi-qubit states or measurements along the eigenbases of multi-qubit observables. This is important in order not to hide some of the complexity of the algorithm realized by the quantum circuit in the initial preparation or the final measurement. With this definition, the complexity of a quantum algorithm can be related to the *size* of its quantum circuit, i.e., the minimal number of gates from \mathcal{G} required to realize the algorithm.

The finiteness property of \mathcal{G} is emphasized because unitary transformations form a continuum. It could be possible to allow \mathcal{G} to contain, e.g., all single-qubit rotations, $U_{\vec{n}}(\theta)$, $\forall \vec{n} \in \mathbb{R}^3$, $\forall \theta \in \mathbb{R}$. However, when we consider different physical implementations of quantum computation, it is more natural to only demand that the same finite set of unitary transformations is realized in all of them. More importantly, as we will discuss in the last chapter, fault-tolerant implementations are possible for gates in a finite set but not for gates in a continuous set.

Due to the continuum of unitary operations, the universality property of our gate set, \mathcal{G} , is more subtle than in the classical case: The most general gate we can apply to the n initial qubits is a unitary transformation in the unitary group in 2^n dimensions, $U(2^n)$. But of course, since our gate set is finite, there will be gates in $U(2^n)$ which we will be unable to implement *exactly* using a finite sequence of gates in \mathcal{G} . However, as we will discuss in §1.2.5, there exist finite gate sets for which any unitary in $U(2^n)$ can be approximated to any desired accuracy using a quantum circuit with a finite number of gates from \mathcal{G} . It is such finite gate sets that we will call *universal* for quantum computation.

Quantum circuits can be represented by space-time diagrams. In these diagrams, time usually progresses from left to right, lines correspond to the “world lines” of qubits and quantum gates are denoted by boxes applied on a subset of these qubits (although, as we will see next, for some gates we use a special notation). The notation for the most common operations in quantum circuit diagrams is shown in figure 1.1, and figure 1.2 gives an example of a quantum circuit. Apart from the Pauli operators in equation (1.4), other common single-qubit rotations are the Hadamard gate,

$$H \equiv \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad (1.9)$$

the phase gate,

$$S \equiv \exp\left(-i\frac{\pi}{4}Z\right) = e^{-i\frac{\pi}{4}} \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}, \quad (1.10)$$

and the $\pi/8$ -gate, $T \equiv S^{1/2}$. We will also often use the two-qubit controlled-NOT gate, CNOT, which

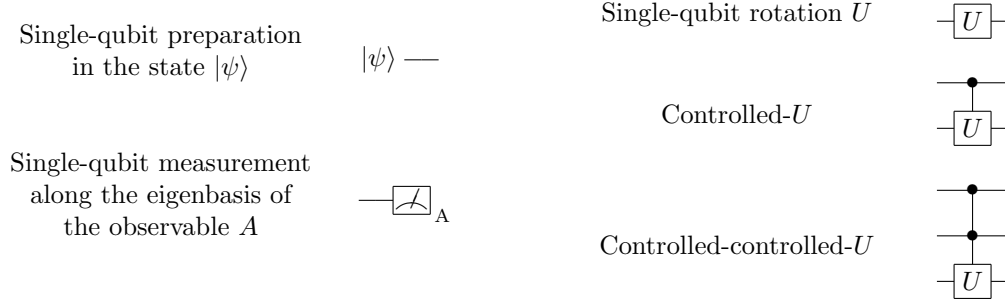


Figure 1.1. Notation for the most common quantum operations. Controlled- U applies U on the second, *target*, qubit when the first, *control*, qubit is in the state $|1\rangle$, otherwise it acts as the identity. Similarly, controlled-controlled- U applies U on the third, *target*, qubit when the first two, *control*, qubits are *both* in the state $|1\rangle$, otherwise it acts as the identity.

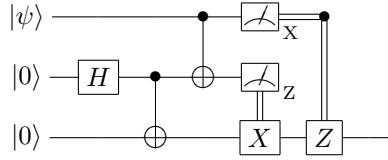





Figure 1.2. An example of a quantum circuit; the direction of time is from left to right. Here, one qubit in the state $|\psi\rangle$ and two ancillary qubits initialized in the state $|0\rangle$ interact by a sequence of Hadamard and CNOT gates. Finally, the first qubit is measured in the eigenbasis of the operator X and the second in the computation basis. Conditioned on the measured eigenvalues being -1 , corrective gates are then applied on the third qubit. It can be verified that the state of the third qubit at the output is $|\psi\rangle$; i.e., the circuit *teleports* the state of the first qubit to the third qubit [27].

acts in the computation basis as

$$\text{CNOT} : |a, b\rangle \rightarrow |a, a + b\rangle, \forall a, b \in \mathbb{Z}_2, \quad (1.11)$$

and whose symbol is , and the controlled- Z gate, CPHASE, which acts in the computation basis as

$$\text{CPHASE} : |a, b\rangle \rightarrow (-1)^{a \cdot b} |a, b\rangle, \forall a, b \in \mathbb{Z}_2, \quad (1.12)$$

and whose symbol is . Finally, we will use the three-qubit controlled-controlled-NOT or Toffoli gate, $\Lambda^2(X)$, whose action in the computation basis was given previously and whose symbol is .

1.2.4 Accuracy

In any physical implementation of a quantum computation, the elementary quantum gates will be executed with some limited accuracy; e.g., the rotation angle in a beam splitter or the length of a pulse can be specified only to some finite precision. In order for the quantum circuit model to

be realistic, it is important to show that any desired accuracy for the computation outcome can be achieved if the accuracy of each gate in the quantum circuit scales at most as an inverse polynomial in the size, L , of the computation. Equivalently, the number of bits of precision specifying the matrix elements of each quantum gate will need to scale “slowly” as at most an inverse polynomial in the *logarithm* of L [28]. In particular, quantum computers would resemble analog rather than digital devices if the number of bits of precision scaled as an inverse polynomial in L , since then physical parameters such as rotation angles or pulse lengths determining the accuracy of each quantum gate would have to be specified to precision that increased as an exponential in L .

Consider an ideal quantum circuit where the sequence of L gates U_1, U_2, \dots, U_L is applied to the initial n -qubit state $|\psi_0\rangle$ where, e.g., $|\psi_0\rangle = |0\rangle^{\otimes n}$. The outcome of the quantum computation is given by a measurement along the computation basis of a subset of the output qubits. Let $\{q_i^{(\text{ideal})}\}$ denote the ideal probability distribution for the measurement outcomes if all gates in the quantum circuit were executed ideally, and let $\{q_i^{(\text{actual})}\}$ denote the actual probability distribution for the measurement outcomes when gates are executed with some limited accuracy. We can then define the *error*, δ , of the actual computation as the L^1 distance between these two probability distributions,

$$\delta = \|q^{(\text{actual})} - q^{(\text{ideal})}\|_1 \equiv \sum_i |q_i^{(\text{actual})} - q_i^{(\text{ideal})}|, \quad (1.13)$$

and let us note the useful property that $\|q^{(1)} - q^{(2)}\|_1 \leq 2$, $\forall q^{(1)}, q^{(2)}$ probability distributions. The *accuracy* of the actual computation will then correspond to $1 - \delta$.

A Stochastic Noise Model

If noise is stochastic, then we can associate probabilities with the occurrence of faults during the computation. Consider the set \mathcal{I}_r of r specific gates, and let us assume that the probability of faults occurring at all gates in \mathcal{I}_r is at most p^r for some $0 \leq p \leq 1$. For example, this would be the case if faults in each gate are described by an i.i.d. process according to which the gate is implemented ideally with probability at least $1 - p$ and it is replaced by a faulty gate with probability at most p . If we denote the probability of having at least one faulty gate in the quantum circuit by $P_{\text{fail}} \leq Lp$,

$$q_i^{(\text{actual})} = (1 - P_{\text{fail}}) q_i^{(\text{ideal})} + P_{\text{fail}} q_i^{(\text{fail})}, \quad (1.14)$$

for some distribution, $\{q_i^{(\text{fail})}\}$, for the computation outcome when faults have occurred. Then, we can upper bound

$$\delta = P_{\text{fail}} \sum_i |q_i^{(\text{fail})} - q_i^{(\text{ideal})}| \leq 2Lp. \quad (1.15)$$

In other words, some desired final accuracy, $1 - \delta_0$, can be achieved if $p \leq \delta_0/2L$.

A Unitary Noise Model

In many physical situations of interest, there exist however noise processes which are not stochastic; e.g., systematic imperfections in the implementation of the gate U_j could lead to implementing a different gate \tilde{U}_j instead. In such a noise model we cannot associate a probability with the event that a given gate is faulty. Instead, the error, ε , in the implementation, \tilde{U}_j , of the ideal gate, U_j , can be defined using a suitable operator norm. Let us assume that for all gates,

$$\|\tilde{U}_j - U_j\|_{\text{sup}} \equiv \sup_{\|\phi\|=1} \|(\tilde{U}_j - U_j)|\phi\rangle\| \leq \varepsilon, \quad (1.16)$$

where the *sup* operator norm is unity for unitary operators and has the properties

$$\|A + B\|_{\text{sup}} \leq \|A\|_{\text{sup}} + \|B\|_{\text{sup}}; \|A \cdot B\|_{\text{sup}} \leq \|A\|_{\text{sup}} \cdot \|B\|_{\text{sup}}; \|A \otimes B\|_{\text{sup}} = \|A\|_{\text{sup}} \cdot \|B\|_{\text{sup}}. \quad (1.17)$$

We first note that the ideal probability distribution of the measurement outcomes is given in terms of the ideal final state $|\psi_{\text{final}}\rangle = U_L \cdots U_2 U_1 |\psi_0\rangle$ as

$$p_i^{(\text{ideal})} = \|\langle i | \psi_{\text{final}} \rangle\|^2 = \text{Tr} \left(P_i \rho^{(\text{ideal})} \right), \quad (1.18)$$

where $P_i \equiv |i\rangle\langle i|$ is the projector onto the final state associated with the outcome i , and $\rho^{(\text{ideal})} \equiv |\psi_{\text{final}}\rangle\langle\psi_{\text{final}}|$. Similarly, if $|\tilde{\psi}_{\text{final}}\rangle = \tilde{U}_L \cdots \tilde{U}_2 \tilde{U}_1 |\psi_0\rangle$ is the actual final state,

$$p_i^{(\text{actual})} = \|\langle i | \tilde{\psi}_{\text{final}} \rangle\|^2 = \text{Tr} \left(P_i \rho^{(\text{actual})} \right), \quad (1.19)$$

where $\rho^{(\text{actual})} \equiv |\tilde{\psi}_{\text{final}}\rangle\langle\tilde{\psi}_{\text{final}}|$. By substituting in equation (1.13),

$$\delta = \sum_i |\text{Tr} \left(P_i (\rho^{(\text{actual})} - \rho^{(\text{ideal})}) \right)| \leq \sum_{i,j} |\lambda_j| \cdot \langle \lambda_j | P_i | \lambda_j \rangle \equiv \|\rho^{(\text{actual})} - \rho^{(\text{ideal})}\|_{\text{tr}}, \quad (1.20)$$

where $\{\lambda_j\}$ are the eigenvalues of $\rho^{(\text{actual})} - \rho^{(\text{ideal})}$ and $\{|\lambda_j\rangle\}$ are the corresponding eigenvectors. Here, $\|\cdot\|_{\text{tr}}$ denotes the *trace norm*, $\|A\|_{\text{tr}} \equiv \text{Tr} \left(\sqrt{A^\dagger A} \right)$, $\forall A \in \mathcal{B}(\mathcal{H}_S)$, and in the last step we have used the completeness relation $\sum_i P_i = I$. By computing the trace norm in equation (1.20),

$$\|\rho^{(\text{actual})} - \rho^{(\text{ideal})}\|_{\text{tr}} = 2\| |\tilde{\psi}_{\text{final}}\rangle - |\psi_{\text{final}}\rangle \| \leq 2\| \tilde{U}_L \cdots \tilde{U}_2 \tilde{U}_1 - U_L \cdots U_2 U_1 \|_{\text{sup}}, \quad (1.21)$$

and finally, by using the triangle inequality of the sup norm and equation (1.16),

$$\| \tilde{U}_L \cdots \tilde{U}_2 \tilde{U}_1 - U_L \cdots U_2 U_1 \|_{\text{sup}} \leq L\varepsilon, \quad (1.22)$$

which implies $\delta \leq 2L\varepsilon$ [28]. Therefore, we arrive at the same conclusion as for stochastic noise,

except that here ε is a bound on the sup norm distance between the actual and ideal gates instead of a probability.

A Non-Markovian Noise Model

The situation we just considered corresponds to the case when the actual noisy gates are unitary operators as, e.g., happens when gate inaccuracies are due to systematic over or under rotations. Let us now generalize this noise model by considering *unitary* faults acting between the computer qubits with Hilbert space \mathcal{H}_S and a common *bath* system with Hilbert space \mathcal{H}_B ; here, the bath is a model for any physically relevant degrees of freedom that have an unwanted coupling to our computer qubits.

Our ideal gates will now correspond to the unitaries $\{U_j \otimes I_B\}$, i.e., they implement the desired unitary, U_j , in \mathcal{H}_S and act trivially in \mathcal{H}_B . Furthermore, let us assume that the actual noisy implementation of the ideal gate $U_j \otimes I_B$ is the unitary $\Delta_j(U_j \otimes I_B)$ where Δ_j acts on the support of U_j and on \mathcal{H}_B . By expanding in the Pauli basis, we may write $\Delta_j = (I \otimes V_j) + F_j$; here, the *fault operator* or, simply, *fault* F_j acts nontrivially in the support of U_j and in some way in \mathcal{H}_B , and V_j , F_j are subject to the constraint that Δ_j is unitary (e.g., V_j may describe the *self-evolution* of the bath). In analogy to assumption (1.16), let us also assume that for all gates,

$$\|F_j(U_j \otimes I_B)\|_{\text{sup}} = \|F_j\|_{\text{sup}} \leq \varepsilon. \quad (1.23)$$

In this noise model, the final state of the computation will be

$$|\tilde{\psi}_{\text{final}}\rangle = \Delta_L(U_L \otimes I_B) \cdots \Delta_2(U_2 \otimes I_B) \cdot \Delta_1(U_1 \otimes I_B) |\psi_0\rangle_{SB}, \quad (1.24)$$

where $|\psi_0\rangle_{SB} = |\psi_0\rangle_S \otimes |\phi_0\rangle_B$ and $|\phi_0\rangle_B$ is arbitrary. Consider next expressing the sequence of unitaries in equation (1.24) as a sum of operators by rewriting each Δ_j as $I \otimes V_j$ plus the fault F_j and opening the parentheses. Let us call each operator in this sum a *fault path*, and consider writing this sum as a sum of two terms: The first “good” term, Gd , contains a single fault path which corresponds to the sequence of the unitaries $\{U_j \otimes V_j\}$. The second “bad” term, Bd , is a sum of all remaining fault paths. This allows us to write $|\tilde{\psi}_{\text{final}}\rangle = (Gd + Bd)|\psi_0\rangle_{SB}$, where $|\psi_{\text{final}}\rangle = Gd|\psi_0\rangle_{SB}$ produces the ideal outcome statistics. Therefore, $\| |\tilde{\psi}_{\text{final}}\rangle - |\psi_{\text{final}}\rangle \| = \| Bd(|\psi_0\rangle_S \otimes |\phi_0\rangle_B) \| \leq \| Bd \|_{\text{sup}}$. Finally, by regrouping the different fault paths in Bd , we rewrite Bd as a sum of terms according to the appearance of the first fault (using some arbitrary time-ordering convention); i.e., in each term, some faulty gate is specified, all gates preceding it are ideal and we sum over fault paths that have or do not have faults in all subsequent gates. There are L terms of this form, each of norm at most ε . Thus, by upper bounding the norm of the sum by the sum of the norms, we obtain $\| Bd \|_{\text{sup}} \leq L\varepsilon$, which implies $\delta \leq 2L\varepsilon$ as before.

1.2.5 Quantum Universality

As in the case of classical computation, the notion of quantum universality is important for two reasons. First, expressing a quantum algorithm as a quantum circuit that uses gates from a given gate set allows us to classify its complexity by means of the size of the quantum circuit. Second, when we consider noisy quantum computation, it becomes sufficient to construct fault-tolerant implementations for all gates in the universal gate set: when a specific algorithm needs to be realized fault-tolerantly, we first express the algorithm in terms of gates in the universal gate set and, then, we implement each gate in this set fault-tolerantly.

Let us start by giving the formal definition of finite quantum universal gate sets.

Definition 1 (Finite Quantum Universal Gate Sets). *Let a finite set of gates $\mathcal{G} = \{g_i\}$. Then, \mathcal{G} is universal for quantum computation if $\forall n > 0, \forall U \in SU(2^n), \forall \varepsilon > 0, \exists$ a finite sequence g_1, g_2, \dots, g_L such that $\|U - g_L \cdots g_2 g_1\|_{\text{sup}} \leq \varepsilon$.*

In other words, the group generated by the gates in \mathcal{G} is *dense* in $SU(2^n)$ (we have used $SU(2^n)$ instead of $U(2^n)$ since overall factors are irrelevant for quantum computation).

There are several finite universal gate sets which are known to admit a fault-tolerant implementation, and some of them may be preferable to others depending on the particular physical implementation of quantum gates. The following proposition lists the two finite universal gate sets that are most commonly used in fault-tolerant quantum computation.

Proposition 1. *The following gate sets are universal for quantum computation:*

1. *Boykin et al. set [29]: $\mathcal{G} = \{\text{CNOT}, H, T\}$,*
2. *Kitaev set [30]: $\mathcal{G} = \{\text{CNOT}, H, S, \text{Toffoli}\}$.*

Another interesting finite universal gate set is the two-element set $\mathcal{G} = \{H, \text{Toffoli}\}$ [31], although it is not particularly useful for fault-tolerant quantum computation.

At this point, it is convenient to introduce the following useful classification of quantum gates.

Definition 2 (C_k Gate Hierarchy [32]). *Let C_1 denote the Pauli group on n qubits which is generated by tensor products of the operators in equation (1.4) and the identity operator with overall phases ± 1 or $\pm i$. Then, for $k > 1$, we recursively define*

$$C_k = \{U|UC_1U^\dagger \subseteq C_{k-1}\}.$$

In particular, C_2 is called the Clifford group and contains operators conjugating Pauli operators to Pauli operators.

The Clifford group, C_2 , is generated by the operators CNOT, H and S —we write $C_2 = \langle \text{CNOT}, H, S \rangle$ where $\langle \cdot \rangle$ denotes a generating set. The C_k hierarchy has the property that, $\forall k$, $C_k \subseteq C_{k+1}$ and $C_{k+1} \setminus C_k \neq \emptyset$. In fact, any gate set that generates the Clifford group and contains one additional gate outside the Clifford group is universal for quantum computation [33]. Henceforth, for simplicity, we will call gates not in C_2 *non-Clifford gates*. The first gate set in proposition 1 contains the non-Clifford gate T ; the second gate set contains the non-Clifford Toffoli gate.

An important question is how efficient is the approximation of a given unitary within a desired error ε , i.e., how many gates from a finite universal set, \mathcal{G} , are necessary to achieve such an approximation. In addition, it is important to know how efficient it is to *find* the gates from \mathcal{G} that realize the desired approximation with a classical computer: if we say that the classical computer *compiles* the gate $U \in SU(2^n)$ in terms of gates in \mathcal{G} , we need to ask how efficient in time and memory is this compilation as a function of ε . The Solovay-Kitaev theorem answers this question.

Theorem 2 (Solovay-Kitaev Theorem [25, 34]). *Consider any fixed number of qubits, n , and let $\mathcal{G} = \{g_i\} \subseteq SU(2^n)$ be a quantum universal gate set that is closed under inversion (i.e., if $g \in \mathcal{G}$ then $g^{-1} \in \mathcal{G}$). Then, $\forall U \in SU(2^n)$, $\forall \varepsilon > 0$, $\forall c > 0$, there exists a sequence g_1, g_2, \dots, g_L such that $\|U - g_L \cdots g_2 g_1\|_{\text{sup}} \leq \varepsilon$, where $L = O\left((\log(1/\varepsilon))^{3+c}\right)$ and the classical time for computing the sequence is $T = O\left((\log(1/\varepsilon))^{3+c}\right)$.*

Thus, the Solovay-Kitaev theorem establishes that gates from \mathcal{G} can efficiently approximate any $U \in SU(2^n)$ within any desired error ε —the efficiency of the approximation is polylogarithmic in $1/\varepsilon$. In addition, this approximation can be found efficiently with a classical computer—the classical running time is also polylogarithmic in $1/\varepsilon$.

Of course, the efficiency of the approximation is in general exponential in the number of qubits, i.e., approximating some $U \in SU(2^n)$ with gates from \mathcal{G} requires quantum circuits of size scaling exponentially in n . But this should not confuse us: The importance of the Solovay-Kitaev theorem is that it implies that any given quantum circuit, \mathcal{C} , which approximates some unitary to some accuracy can be efficiently converted to another quantum circuit, \mathcal{C}' , which achieves an improved accuracy—this conversion is efficient since it is achieved by approximating each gate in \mathcal{C} which only acts on a fixed constant number of qubits.

1.3 Quantum Coding

In this section, I will discuss how quantum information can be encoded in such a way that errors which only affect a sufficiently “small part” of it can first be digitized and then corrected. For this discussion, I will assume that the error-correction procedure that diagnoses and corrects errors is realized itself ideally without faults. The more general question of how quantum error correction

can be performed with noisy operations is the subject of the theory of fault-tolerant quantum computation and will be discussed in detail in the main body of this thesis.

In the classical setting, the idea of encoding information in a way that enables error correction is intuitively clear. The most natural classical encoding scheme encodes information by repeating it several times. For example, when we want to transmit the bit value 0 or 1 over a noisy transmission line, we may send the message several times as a string of 0s or 1s so that the receiver will have a high probability of deducing the intended message by taking a majority vote. The theory of classical error correction includes much more general and efficient schemes than this simple encoding by repetition. However, the principal idea is the same in all these encoding schemes: In order to protect classical information from errors, information is encoded using some form of *redundancy* so that if a sufficiently small part of it is affected by errors, the receiver can recover the correct transmitted information with high probability.

Protecting quantum information from errors is far more challenging due to the *information-disturbance principle*. Put simply, this principle is just the statement that when we gain information about an unknown quantum state we unavoidably disturb it: Since quantum states cannot be copied [35], the only possible way to gain information about an unknown quantum state is to perform a measurement on it. But then, the post-measurement state is an eigenstate of the measured operator—as a consequence of the measurement, the initial quantum state *collapses* to a post-measurement state which is, in general, different from the initial one. Therefore, quantum error correction procedures must be able to diagnose errors without gaining any information about the encoded quantum states, since if they gain any such information, the encoding will be damaged.

Despite this obstacle, quantum codes exist [36, 37]. Their construction is based on the following two general ideas: First, quantum codes achieve redundancy not by repeating quantum information in different quantum system but by encoding information into quantum states that are *entangled* across many different quantum systems. Second, errors in these entangled encoded quantum states can be diagnosed by making multi-qubit measurements which do not reveal information about the encoded information. In essence, quantum codes operate by encoding quantum information into the *quantum correlations* between different systems. Intuitively, the idea is that an adversary that can only apply errors on some small fraction of these systems will be unable to gain information about the encoded state since information is hidden in the correlations across a large number of different systems—we may say that information is encoded “non-locally” and it is protected as long as errors only act “locally.”

The rest of this section is structured as follows: In §1.3.1, I discuss the basic principles of quantum coding and I state the criteria for quantum error correction. In §1.3.2, I review the stabilizer formalism that provides a succinct description of a wide class of quantum codes. Next, in §1.3.3, I discuss the CSS class of quantum codes which are well-suited for fault-tolerant quantum

computation. Finally, in §1.3.4, I present two examples of quantum codes.

1.3.1 Quantum Error-Correction Criteria

Consider a *system* with a Hilbert space \mathcal{H}_S that includes a natural decomposition into a tensor product of, say, n two-dimensional subsystems which will be the qubits of our quantum computer. According to theorem 1, any physical noise operation, \mathcal{N} , acting on \mathcal{H}_S can be described in terms of its Kraus operators, $\{M_k\}$, which are subject to the constraint $\sum_k M_k^\dagger M_k = I_S$. In particular, each operator M_k can be seen as arising due to a unitary evolution in a larger Hilbert space, $\mathcal{H}_S \otimes \mathcal{H}_B$, that includes a *bath* or *environment*, B , followed by a measurement in the bath along the orthonormal basis $\{|k\rangle_B\}$ that gave the outcome k . Since quantum measurement is an irreversible process, interacting the system and the bath and then performing a measurement on the bath will, in general, result in an irreversible change in the state of the system. Reversing the action of noise would require reversing the action of the Kraus operators which is only possible if all but one are zero (in which case the non-zero Kraus operator would be unitary); we can state this simply by saying that the only physical operations which are invertible correspond to unitary operations. We cannot therefore hope to be able to invert an arbitrary physical noise operation acting on all the qubits of our quantum computer. In the absence of more information about the structure of noise, the goal of quantum error correction is more modest: the goal is to design a correction procedure for protecting against noise that acts on sufficiently small subsets of qubits.

Let the initial state in \mathcal{H}_S be the vector $|\psi\rangle_S$. (In the more general case where the state is described by a density matrix, ρ , we may always expand $\rho = \sum_i c_i |\psi_i\rangle\langle\psi_i|_S$, e.g., in terms of the eigenvectors $\{|\psi_i\rangle_S\}$ of ρ , and we can consider the vector $|\psi\rangle = \sum_i \sqrt{c_i} |\psi_i\rangle_S \otimes |i\rangle_R$ where $\{|i\rangle_R\}$ is an orthonormal basis in a virtual *reference* system with Hilbert space \mathcal{H}_R which has been added to \mathcal{H}_S . The idea is that if we later trace over the reference system, then we are left with ρ , i.e., $\text{Tr}_R(|\psi\rangle\langle\psi|) = \rho$. We call this process, by which a density matrix, ρ , is seen as arising from a state vector, $|\psi\rangle$, in a larger space, *purification* of ρ .) By theorem 1, we can describe \mathcal{N} as performing the operation

$$|\psi\rangle_S \otimes |0\rangle_B \rightarrow \sum_k M_k |\psi\rangle_S \otimes |k\rangle_B, \quad (1.25)$$

where each Kraus operator, M_k , may be expanded in a fixed basis of operators, $\{E_a\}$, that we will call the *error* basis. Then, the action of \mathcal{N} becomes

$$|\psi\rangle_S \otimes |0\rangle_B \rightarrow \sum_a E_a |\psi\rangle_S \otimes |a\rangle_B, \quad (1.26)$$

where the states $\{|a\rangle_B\}$ in \mathcal{H}_B are, in general, not normalized nor orthogonal. If $\{|a\rangle_B\}$ were in fact orthogonal and normalized then the operators $\{E_a\}$ could be interpreted as different errors which

have occurred on $|\psi\rangle_S$ depending on the state in \mathcal{H}_B . Even though this interpretation is not valid in general, it is sufficient to proceed by devising a correction procedure that inverts each E_a separately. Our correction procedure is implemented by a physical recovery operation, \mathcal{R} , with Kraus operators $\{R_\mu\}$; then, the composition $\mathcal{R} \circ \mathcal{E}$ acts on $|\psi\rangle_S$ as

$$|\psi\rangle_S \otimes |0\rangle_B \otimes |e_0\rangle_A \rightarrow \sum_{a,\mu} R_\mu E_a |\psi\rangle_S \otimes |a\rangle_B \otimes |e_\mu\rangle_A, \quad (1.27)$$

where $\{|e_\mu\rangle_A\}$ is an orthonormal basis in an ancillary system with Hilbert space \mathcal{H}_A initialized it in the state $|e_0\rangle_A$ which we attach to \mathcal{H}_S in order to implement the recovery.

Error recovery is successful if the state on the right-hand side of equation (1.27) is of the form $|\psi\rangle_S \otimes |\phi\rangle_{BA}$, for any state $|\phi\rangle_{BA} \in \mathcal{H}_B \otimes \mathcal{H}_A$. We have already discussed that this is not always possible for all states, $|\psi\rangle_S$, and all noise operations, \mathcal{N} . However, successful error recovery is possible when $|\psi\rangle_S$ is restricted in a *subspace* of \mathcal{H}_S and \mathcal{N} is restricted to act nontrivially on a sufficiently small subset of the n qubits in \mathcal{H}_S . The following theorem states the criteria for when quantum error correction is possible.

Theorem 3 (Quantum Error-Correction Criteria [38, 39]). *Let an error basis, $\{E_a\}$, and let P_L be the projector onto a subspace, \mathcal{H}_L , of \mathcal{H}_S . Then the following are equivalent:*

1. *The noise operation, \mathcal{N} , whose Kraus operators are linear combinations of the $\{E_a\}$ with arbitrary complex coefficients is correctable, i.e., there exists an ancillary system with Hilbert space \mathcal{H}_A that is initialized in the state $|e_0\rangle_A$ and a recovery operation, \mathcal{R} , acting on $\mathcal{H}_S \otimes \mathcal{H}_A$ such that, $\forall \rho_L \in \mathcal{H}_L$, $\text{Tr}_A (\mathcal{R}(\mathcal{E}(\rho_L) \otimes |e_0\rangle\langle e_0|_A)) = \rho_L$,*
2. *$\forall E_a, \forall E_b, \exists$ a Hermitian complex matrix, c , such that $P_L E_a^\dagger E_b P_L = c_{ab} P_L$.*

Quantum information will be encoded in states $|\psi\rangle_L \in \mathcal{H}_L$; we will call \mathcal{H}_L the *code space* and $|\psi\rangle_L$ the *logical* $|\psi\rangle$ state. Then, the condition $P_L E_a^\dagger E_b P_L = c_{ab} P_L$ can be understood as saying that (i) two orthogonal logical states must remain orthogonal after errors act on them (so that errors do not make logical states which were initially orthogonal to later overlap), and (ii) different logical states must be transformed by different errors in a way that only depends on the errors and not on the logical states themselves (so that learning about the errors will not give us any information about the logical state).

We note that due to the linearity of quantum mechanics, if a code can correct all errors in the basis of Pauli operators on up to t qubits, then it can also correct *any* operator, M_k , supported on up to t qubits (since the Pauli operators form a basis in the operator space). It follows that any noise operation with support on up to t qubits can also be inverted. Therefore, in most cases of interest it is sufficient to design codes that correct Pauli errors such as the stabilizer codes that will be discussed in the next section; henceforth, we will only consider such codes.

1.3.2 Stabilizer Codes

Let us now briefly review the theory of binary stabilizer codes [40, 41, 42] that form a particularly interesting subclass of quantum codes. The code space of a stabilizer code is the *simultaneous* eigenspace of a normal abelian subgroup, \mathcal{S} , of C_1 that does not contain $-I$ or $\pm iI$; this subgroup is called the code's *stabilizer* since, by convention, $\forall S \in \mathcal{S}$, $S|\psi\rangle_L = |\psi\rangle_L$.

To understand the error-correction capabilities of a stabilizer code, a useful concept is the code's distance.

Definition 3 (Distance). *Let the weight of the Pauli operator E_a , $\text{wgt}(E_a)$, be the number of qubits in its support (i.e., the number on qubits on which E_a acts different than the identity). Then, with the same notation as in theorem 3, the distance, d , of a code is $d = \min\{\text{wgt}(E_a) | P_L E_a P_L \not\propto P_L\}$.*

In other words, the distance is the smallest number of qubits on which an error, E_a , needs to act nontrivially before the condition $P_L E_a P_L \propto P_L$ is violated. This implies that $P_L E_a^\dagger E_b P_L \propto P_L$ is satisfied for all $E_a^\dagger E_b$ that act nontrivially on up to $d - 1$ qubits, or that the code can correct all errors with support on up to $t = \lfloor \frac{d-1}{2} \rfloor$ qubits. Assuming the dimension of \mathcal{H}_L is 2^k , such a code will be denoted as an $[[n, k, d]]$ code—the notation means that n physical qubits are used to encode k logical qubits and the distance of the code is d .

The analysis of stabilizer codes is facilitated by the simple algebra of the Pauli group. Let us define the *centralizer* of \mathcal{S} in C_1 by $C(\mathcal{S}) = \{E \in C_1 | \forall S \in \mathcal{S}, [E, S] = 0\}$; i.e., $C(\mathcal{S})$ is the set of all Pauli operators that commute with every element of the code's stabilizer. Because of the special properties of the Pauli group, the centralizer equals the *normalizer*, $N(\mathcal{S})$, of \mathcal{S} in C_1 , i.e., the set of Pauli operators that leave \mathcal{S} invariant under conjugation. Clearly, \mathcal{S} is a normal subgroup of $N(\mathcal{S})$, and $N(\mathcal{S})/\mathcal{S}$ is a group of order 2^{2k} containing the four logical Pauli operators for each of the k logical qubits. Then, $\forall E_a, E_b \in C_1$, either (i) $E_a^\dagger E_b \in \mathcal{S}$, in which case $P_L E_a^\dagger E_b P_L = P_L$, or (ii) $E_a^\dagger E_b \in N(\mathcal{S}) \setminus \mathcal{S}$, in which case $P_L E_a^\dagger E_b P_L = \mathcal{O}_{ab} P_L$ where \mathcal{O}_{ab} is a nontrivial logical Pauli operator, or (iii) $\exists S \in \mathcal{S}$ such that $S E_a^\dagger E_b = -E_a^\dagger E_b S$, which implies $P_L E_a^\dagger E_b P_L = 0$. Since the error-correction criteria fail only in case (ii), it follows that the distance, d , of the code is the minimal weight among all elements in $N(\mathcal{S}) \setminus \mathcal{S}$, i.e., the minimal weight among all logical Pauli operators.

A more succinct description of stabilizer codes is possible if we represent Pauli operators as binary vectors. The subscript a labelling an operator $E_a \in C_1$ can be seen as a length- $2n$ binary vector, where

$$E_a = \bigotimes_{j=1}^n \left(i^{a[j] \cdot a[j+n]} X^{a[j]} Z^{a[j+n]} \right), \quad (1.28)$$

and $a[j]$ denotes the j th component of a . With this notation, two elements E_a and E_b of C_1 obey the commutation relation

$$E_a E_b = (-1)^{a \Lambda b^T} E_b E_a, \quad (1.29)$$

where T denotes matrix transposition, and

$$\Lambda = \begin{pmatrix} \mathbf{0}_n & \mathbf{I}_n \\ \mathbf{I}_n & \mathbf{0}_n \end{pmatrix}; \quad (1.30)$$

here $\mathbf{0}_n$ and \mathbf{I}_n are the zero and identity $n \times n$ matrices, respectively.

For a stabilizer code with k logical qubits, the generators $\{G_{a_1}, G_{a_2}, \dots, G_{a_{n-k}}\}$ of \mathcal{S} can be represented by a $(n-k) \times 2n$ binary *check* matrix,

$$G \equiv \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_{n-k} \end{pmatrix}. \quad (1.31)$$

Each generator squares to the identity and has eigenvalues ± 1 with equal degeneracy. Therefore, the $n-k$ generators specify 2^{n-k} mutually orthogonal subspaces, each of dimension $2^n/2^{n-k} = 2^k$. In particular, the subspace that is the common $+1$ eigenspace for all generators is the *code space*; it is the subspace where, by convention, we encode our k logical qubits. Each subspace can be labeled by a length- $(n-k)$ binary vector, e , where the eigenvalue of G_{a_i} is $(-1)^{e[i]}$, for $i = 1, \dots, n-k$. This vector e is called the *syndrome* of the subspace; specifically, the code space has syndrome $e = (00 \dots 0)$. Furthermore, equation (1.29) implies that the action of a Pauli operator E_a on the code space changes the syndrome to the value $e = a\Lambda G^T$.

For a *non-degenerate* stabilizer code that corrects $t = \lfloor \frac{d-1}{2} \rfloor$ errors, all $E_a \in C_1$ of weight at most t take the code space to mutually orthogonal subspaces with distinct syndromes; thus, under the assumption that no more than t errors occurred, every different value of the syndrome points to a unique Pauli error operator. In implementing error correction, the syndrome, e , is first measured, and then E_a^\dagger is applied to invert the error E_a , where E_a is the unique Pauli operator with $\text{wgt}(E_a) \leq t$ such that $e = a\Lambda G^T$. If the code is *degenerate*, then E_a may not be unique but each E_a of weight up to t satisfying $e = a\Lambda G^T$ is equally effective in correcting the error. Indeed, if E_a and E_b give the same syndrome then $E_a^\dagger E_b$ has syndrome zero (i.e., $E_a^\dagger E_b \in N(\mathcal{S})$) and since $\text{wgt}(E_a^\dagger E_b) \leq 2t < d$, it is in fact $E_a^\dagger E_b \in \mathcal{S}$; therefore, applying E_a^\dagger to recover from the error E_b results in a combined operator that acts trivially in the encoded information and, so, recovery is successful. The property of a code being degenerate is therefore consequence of the fact that there exist pairs of distinct errors with identical syndromes whose product is in \mathcal{S} , and has no analogue in classical coding.

Let us finally comment on the case when the syndrome value does not correspond to any correctable error operator—codes for which this never occurs are called *perfect*. If a code is not perfect and a syndrome is obtained that cannot be associated with a correctable error—i.e., if there are

two equal-weight errors E_a and E_b that give the same syndrome and $E_a^\dagger E_b \in N(\mathcal{S}) \setminus \mathcal{S}$ —successful error recovery is not possible. In that case, error correction simply maps its input to the code space, i.e., it applies some recovery operator which restores the syndrome to zero; for the purposes of our discussion, it will not matter what the convention is for how this recovery operator is chosen.

1.3.3 CSS Construction

A subclass of stabilizer codes is CSS codes [43, 44] which are especially well-suited for fault-tolerant quantum computation. CSS codes are constructed from two classical binary linear codes whose theory we will first briefly review.

Letting \mathbb{F}_2^n denote the vector space defined by length- n strings over the binary field \mathbb{F}_2 , a classical binary linear code, $C = [n, k, d]$, is a subspace of \mathbb{F}_2^n , the *code space*, which is spanned by the k length- n binary vectors $\{v_1, v_2, \dots, v_k\}$. The code words are linear combinations of the vectors $\{v_i\}$, i.e., the length- k binary word $w = (w_1, w_2, \dots, w_k)$ is encoded as the code word $w_L \equiv \sum_{i=1}^k w_i v_i$. Code words have the property that $H \cdot w_L^T = 0$, where H is the $(n-k) \times n$ binary *parity check* matrix satisfying $H \cdot v_i^T = 0, \forall v_i$. Given a code word, w_L , which has been affected by an error, Δw_L , error correction proceeds by computing the length- $(n-k)$ binary *syndrome* vector, $e = H \cdot (w_L + \Delta w_L)^T = H \cdot (\Delta w_L)^T$, finding the most likely error, $\tilde{\Delta w}_L$, leading to e and applying $\tilde{\Delta w}_L$ to invert the error, $w_L + \Delta w_L + \tilde{\Delta w}_L = w_L$ with high success probability. Let the *Hamming weight* of a binary string be the number of 1s in it. Then, the code's *distance*, d , is the minimal Hamming weight of a nonzero code word, and all errors of Hamming weight up to $\lfloor \frac{d-1}{2} \rfloor$ can be corrected. Finally, the *dual* code to C , C^\perp , is the orthogonal complement of C in \mathbb{F}_2^n , i.e., all code words of C^\perp are orthogonal to all code words of C (it can be that $C \subseteq C^\perp$ or $C^\perp \subseteq C$).

CSS quantum codes are constructed from two classical binary linear codes $C = [n, k, d]$ and $C' = [n, k', d']$ such that $(C')^\perp \subseteq C$. The construction proceeds by using C with parity check matrix H to correct Z errors and by using C' with parity check matrix H' to correct X errors, i.e., by creating the quantum check matrix,

$$G \equiv \left(\begin{array}{c|c} H_{(n-k) \times n} & \mathbf{0}_{(n-k) \times n} \\ \hline \mathbf{0}_{(n-k') \times n} & H'_{(n-k') \times n} \end{array} \right), \quad (1.32)$$

where subscripts denote the dimensions of the submatrices. The condition $(C')^\perp \subseteq C$ is needed in order for rows of G to commute which requires $H \cdot (H')^T = H' \cdot H^T = 0$. Indeed, this requirement implies that the rows of H' span a code space that is contained in the code space of C (since every code word in C satisfies $H \cdot w_L^T = 0$); but the rows of H' span the code space of $(C')^\perp$ since they are orthogonal to every code word in C' .

According to equation (1.32), each row of H becomes a generator of the code's stabilizer that is

a tensor product of X operators alone (and the identity on the remaining qubits); such generators we will call X -type. Similarly, each row of H' becomes a generator that is a tensor product of Z operators alone; such generators we will call Z -type. Since there are $2n - k - k'$ independent generators, the quantum code encodes $n - (2n - k - k') = k + k' - n$ logical qubits. Furthermore, by measuring the eigenvalues of the X -type (respectively, Z -type) generators and computing the syndrome of the classical code C (respectively, C'), we can correct Z (respectively, X) errors on up to $\lfloor \frac{d-1}{2} \rfloor$ (respectively, $\lfloor \frac{d'-1}{2} \rfloor$) qubits. And since $Y = iXZ$, a Y error will be corrected as being an X and a Z error on the same qubit. Therefore, the CSS code just constructed is an $[[n, k + k' - n, \min(d, d')]]$ code.

1.3.4 Examples

Our first example is a CSS code that is generated from the classical binary $C = [7, 4, 3]$ code which contains its dual code. The parity check matrix of C is

$$H \equiv \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}, \quad (1.33)$$

and its code space is spanned by the three rows of H and the additional code word $(1, 1, 1, 0, 0, 0, 0)$. C^\perp is the even subcode of C and its code space is spanned by the rows of H . Therefore, $C^\perp \subseteq C$ and we may take $C = C' = [7, 4, 3]$ in the CSS construction. This results in Steane's $[[7, 1, 3]]$ code [37] which encodes one logical qubit in a *block* of seven physical qubits and corrects any noise operation whose Kraus operators have support on at most one physical qubit—we will often simply say that the code corrects any single error.

Our second example is Shor's code [36]. This code is based on the classical repetition code, $C_Z^{(n)}$, which encodes one logical bit into a string of n bits via the encoding $w \rightarrow w_L = (w, w, \dots, w)$. Its parity check matrix is

$$H \equiv \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & \dots \\ 0 & 1 & 1 & 0 & 0 & \dots \\ 0 & 0 & 1 & 1 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \quad (1.34)$$

Rephrasing $C_Z^{(n)}$ as a quantum stabilizer code, its check matrix is

$$G(C_Z^{(n)}) \equiv \left(0 \mid H \right), \quad (1.35)$$

and it follows that it corrects X errors on up to $\lfloor \frac{n-1}{2} \rfloor$ qubits. Its logical X operator that exchanges

$|0\rangle_L$ and $|1\rangle_L$ is $X_L = X \otimes X \otimes X \otimes \cdots$, i.e., an X , or NOT, operator acting on all bits. Its logical Z operator that multiplies the phase of $|1\rangle_L$ by -1 and leaves $|0\rangle_L$ unchanged is, e.g., $Z_L = Z \otimes I \otimes I \otimes \cdots$, which shows that the code cannot correct any Z errors.

If we interchange the X and Z operators, we can consider the similar code, $C_X^{(n)}$, with parity check matrix

$$G(C_X^{(n)}) \equiv \left(H \mid 0 \right), \quad (1.36)$$

which corrects Z errors on up to $\lfloor \frac{n-1}{2} \rfloor$ qubits. The logical Z operator of this code is $Z \otimes Z \otimes Z \otimes \cdots$ and the logical X operator is $X \otimes I \otimes I \otimes \cdots$, which implies that $C_X^{(n)}$ corrects no X errors.

Shor's code is the *concatenated* $C_X^{(n)} \circ C_Z^{(n)}$: We partition n^2 qubits into n subblocks of n qubits each. Within each subblock we use $C_Z^{(n)}$ to correct X errors. The idea of *concatenation* is to further encode the n logical qubits of the n subblocks by using the “level-2” code $C_X^{(n)}$; this results in one level-2 logical qubit. Its logical X operator is $X_L \otimes I \otimes \cdots$, where the tensor product structure denotes the n different subblocks. Similarly, its logical Z operator is $Z_L \otimes Z_L \otimes Z_L \otimes \cdots$, i.e., a Z operator applied to the first qubit in each of the n subblocks. Since the weight of these logical X and Z operators is n , Shor's code has distance n and it can correct arbitrary errors on up to $\lfloor \frac{n-1}{2} \rfloor$ qubits; it is an $[[n^2, 1, n]]$ code.

Chapter 2

The Syntax of Fault-Tolerant Quantum Computation

2.1 Introduction

In this chapter, I will describe quantum circuit designs that are used in implementing quantum computation in a way that is resilient to noise—we will say that such circuits are *fault-tolerant*. Similar to the case of fault-tolerant classical computation, the design of fault-tolerant quantum circuits is based on the idea of *simulating* each quantum operation in an ideal quantum circuit using operations encoded in an error-correcting code.

The analysis of fault-tolerant quantum circuit simulations has many common features with fault-tolerant simulations of classical circuits as, e.g., analysed by Von Neumann [45] and by Gács [46, 47]. However, there are two main features in the quantum version of this problem that have no classical analog. First, quantum error correction needs to deal with phase flip (Z) in addition to *classical* bit flip (X) errors, and it must also operate without gaining any information about the encoded information. And second, the fault-tolerant implementation of a universal set of *encoded* (or *logical*) operations is significantly more challenging in the quantum than in the classical case and it requires the off-line preparation and verification of special ancillary quantum states known as *quantum software*.

For the fault-tolerant methods discussed in this thesis to be effective, there are certain assumptions about the experimental setup and the structure of noise that are essential. The following proposition states them.

Proposition 2. *The following assumptions are necessary for our fault-tolerant methods to be effective and for proving the quantum threshold theorem:*

1. *Faults that act collectively on many qubits are highly suppressed in probability or in amplitude depending on the noise model. Furthermore, the noise strength, ε , must be a sufficiently small constant that is independent of the size of the computation.*

2. *There is an inexhaustible supply of fresh ancillary qubits, or the ability to refresh and reuse ancillary qubits an indefinite number of times; otherwise, there will be a limit on the amount of entropy arising from noise that we can flush from the quantum computer [48].*
3. *It is possible to apply quantum gates in parallel on disjoint sets of qubits; otherwise we will be unable to fight against noise in memory that afflicts all parts of the computer simultaneously.*

This chapter is organized as follows: In §2.2, I define the local noise model that will be considered throughout this thesis. Then, in §2.3, I will state the key properties fault-tolerant quantum circuits need to satisfy. These properties will be stated *syntactically* rather than semantically, i.e., as properties of the noisy circuits themselves and independent of the actual quantum state being processed by the quantum computer. This syntactic approach will prove to be extremely useful in proving the basic lemma of this chapter that relates the strength of local noise and the accuracy of the fault-tolerant simulation. We will find that for local noise, fault-tolerant simulations achieve greater accuracy than the unencoded computation provided the noise strength is below a certain critical value. In the next chapter, we will use this result to show that the accuracy of fault-tolerant simulations can be improved to *any* desired level.

2.2 Local Noise

Consider a quantum circuit computation. Elementary quantum operations in this circuit consist of single-qubit preparations, quantum gates chosen from a finite universal gate set (including identity gates that correspond to storage of a qubit in memory), and single-qubit measurements. Let the term *location* denote any of these operations since each operation can be viewed as a point in the space-time quantum circuit diagram; i.e., a location is a label indicating which qubits interact via what quantum operation at what time step of the computation. Due to noise, elementary operations may deviate from their ideal realization resulting in what we may call *faults* during the computation.

Let us start by recalling the non-Markovian noise model we discussed in the introduction. In this noise model, the actual implementation of every ideal gate is a unitary acting on the support of the ideal gate and a common “bath” system, where the difference of this unitary from the ideal gate has small sup norm. The noisy implementation of an ideal gate that applies $U_j \otimes I_B$ to the qubits on the support of U_j and the bath is the *unitary* $\Delta_j(U_j \otimes I_B)$, where $\Delta_j = (I \otimes V_j) + F_j$ and the *fault* F_j acts nontrivially on the support of U_j . In this situation, let us say that the noise has *strength* ε if for all gates,

$$\|F_j\|_{\text{sup}} \leq \varepsilon. \quad (2.1)$$

A noisy measurement can be modeled as Δ_j followed by an ideal measurement, and a noisy preparation of a quantum state can be modeled as an ideal preparation followed by Δ_j . For gates

that are executed in parallel during the *same* time step, a convention is necessary to specify the time ordering of the, possibly non-commuting, operators $\{\Delta_j\}$ acting on the system and the bath; however, the noise strength does not depend on how the time-ordering ambiguity is resolved.

We also recall the definition of the concept of *fault paths*: If some ideal quantum circuit applies the sequence of gates $U_1 \otimes I_B, U_2 \otimes I_B, \dots, U_L \otimes I_B$, its noisy implementation will correspond to the actual unitary $\Delta_L(U_L \otimes I_B) \cdots \Delta_1(U_1 \otimes I_B)$ ¹. By expanding the $\{\Delta_j\}$, we obtain a sum of terms where, for each term, either $I \otimes V_j$ or the fault operator, F_j , has been applied at the noisy location j . Each term in this sum is a fault path and it indicates which locations in the quantum circuit are implemented ideally and which locations are faulty.

For each fault path, we may consider the corresponding operator acting on the bath. If different fault paths are labeled by perfectly distinguishable states of the bath, we expect that probabilities can be associated with faults. More specifically, if for all locations the operators $I \otimes V_j$ and F_j map the bath to orthogonal states, then we expect that different fault paths will not interfere so that our noise model describes faults that are independent and stochastic (in essence, we assume that there exists a mechanism in the bath that effectively projects onto one of these perfectly distinguishable states, thus preventing the interference between different fault paths). In this case, the probability of a fault is at most $p = \varepsilon^2$ —we may regard F_j as the *amplitude* weighting a fault, whose norm square is a probability. If instead $\{F_j\}$ decompose into a tensor-product form between the system and the bath, then our noise model describes unitary faults which do not involve the bath. In general, even if faults involve the bath, it may not be the case that different faults are associated with perfectly distinguishable states of the bath. In that case we cannot associate a well-defined probability with faults at different locations and different fault paths will *interfere* as they can add coherently. Nevertheless, using the properties (1.17), the sup norm of the *sum* of all fault paths with faults at some r *specific* locations is at most ε^r [49]. This fact motivates the following definition.

Definition 4 (Local Noise). *Consider a noisy quantum circuit realized as a unitary transformation acting on a system and a bath. We may express this unitary transformation as a sum of fault paths where, for each fault path, each location is either faulty or ideal. Let \mathcal{I}_r denote a set of r specific locations, and let $F(\mathcal{I}_r)$ denote the sum over all fault paths for which all locations in \mathcal{I}_r are faulty (with the fault paths unrestricted outside of \mathcal{I}_r). Then, noise is local with strength ε if, for all \mathcal{I}_r , $\|F(\mathcal{I}_r)\|_{\text{sup}} \leq \varepsilon^r$. (If faults are stochastic, then noise is local with strength p if, for all \mathcal{I}_r , the probability of $F(\mathcal{I}_r)$ is at most p^r .)*

It should be emphasized that the noise model described in the beginning of this section is a particular example of local noise—there exist other noise models that also satisfy the condition of definition 4 as, e.g., the noise model in [50]. The defining characteristic of local noise is that there is

¹Our operator-ordering convention is that operators on the right act first on the input state.

an exponential penalty for faults that occur collectively at once on many qubits that do not interact in the ideal computation.² However, faults at different locations are allowed to act on a common bath that can store information about faults for an indefinitely long time. Hence, local noise is in general *non-Markovian* and the presence of a common bath allows faults to be correlated both in time and in space. We will see that these correlations do not prevent fault-tolerant error correction from being effective, provided the noise strength is sufficiently weak.

From a microscopic perspective, the physics of local noise is most naturally formulated in terms of a Hamiltonian, H , that describes the joint evolution of the system and the bath. We may generally express H as

$$H = H_S + H_B + H_{SB} , \quad (2.2)$$

where H_S is the time-dependent Hamiltonian of the system that generates the evolution of the ideal quantum computation, H_B is an arbitrary Hamiltonian of the bath, and the time-dependent Hamiltonian H_{SB} couples the system to the bath. In particular, the local noise model described in the beginning of this section can be generated by a coupling of the form

$$H_{SB} = \sum_j H_{SB,j} , \quad (2.3)$$

where each term $H_{SB,j}$ describes how the set of qubits on the support of location j interact with the bath; i.e., at any given time step, different qubits are *directly* coupled by H_{SB} only if the ideal computation Hamiltonian, H_S , also couples them in the same time step. We emphasize that because the qubits at every location j are coupled via $H_{SB,j}$ to a *common* bath, faults may be correlated both spatially and temporally. In particular, these correlations may be chosen adversarially since we have not imposed any locality condition on the Hamiltonian of the bath, H_B , that is completely arbitrary.

We can express the time evolution governed by the Hamiltonian H in terms of *time-resolved* fault paths [49]: In order to study the evolution for total time T , we divide T into N time intervals each of width $\Delta t = T/N$, and take the limit $N \rightarrow \infty$ so that terms of order $(\Delta t)^2$ can be neglected. Then, the evolution operator for the total time interval $(0, T)$ can be Lie-Trotter expanded as

$$U(0, T) = \lim_{N \rightarrow \infty} \prod_{i=1}^N U_S(\Delta t_i) U_{SB}(\Delta t_i) U_B(\Delta t_i) , \quad (2.4)$$

where $U_S(\Delta t_i)$, $U_B(\Delta t_i)$ describe the time evolution during the i th interval in the system and the bath, respectively³. By taking the coupling Hamiltonian, H_{SB} , as in equation (2.3), $U_{SB}(\Delta t_i)$ for

²For example, if \mathcal{I}_2 labels two qubits that do not interact during some time step in the quantum computation, then $F(\mathcal{I}_2) \leq \varepsilon^2$. In contrast, if the same two qubits interact via a CNOT at the same time step, then they are both considered to be part of *one* location, \mathcal{I}_1 , and $F(\mathcal{I}_1) \leq \varepsilon$.

³There is an implicit time ordering in equation (2.4) since the operators acting at different time intervals may not

the i th interval can be expanded as

$$U_{SB}(\Delta t_i) \approx \prod_j (I_{SB} - iH_{SB,j}\Delta t) , \quad (2.5)$$

where j runs over all locations in the time interval $(0, T)$ (i.e., if $(0, T)$ corresponds to one time step in the computation, j runs over all gates that are executed during that time step). By opening the parentheses in equation (2.5), we obtain a sum of fault paths: at each time interval labelled by i , either the identity, I_{SB} , or the *micro-fault*, $-iH_{SB,j}\Delta t$, is inserted at each location j .

Suppose that the maximum time required for implementing an operation in the quantum circuit is τ , so that each location can be viewed as a *coarse graining* of N consecutive time intervals of width at most $\Delta t = \tau/N$. We also observe that a specific coarse-grained location, j , is faulty if there is a micro-fault at any one of the N time intervals it contains. Then, in the time-resolved fault path expansion at location j , the term $\prod_{i=1}^N U_S(\Delta t_i)U_B(\Delta t_i)$ corresponds to the ideal implementation, whereas the sum, F_j , of all other terms corresponds to a faulty implementation.

Let us now calculate an upper bound on $\|F_j\|_{\text{sup}}$. In all terms in the sum F_j , a micro-fault is inserted in at least one of the N time intervals. Therefore, we can rewrite F_j as a sum of N terms where a micro-fault is inserted in some time interval, each preceding time interval has been replaced by I_{SB} and each succeeding time interval has been replaced by the unitary $U_{SB}(\Delta t_i)$. Thus, $\|F_j\|_{\text{sup}} \leq \lim_{N \rightarrow \infty} N\| -iH_{SB,j}\Delta t\|_{\text{sup}} = \tau\|H_{SB,j}\|_{\text{sup}}$. If we now suppose that at all noisy locations the system-bath coupling Hamiltonian satisfies

$$\|H_{SB,j}\|_{\text{sup}} \leq \lambda_0 , \quad (2.6)$$

we can conclude that the norm of a fault acting at a specific coarse-grained location, j , satisfies $\|F_j\|_{\text{sup}} \leq \lambda_0\tau$. This provides a justification in terms of a microscopic Hamiltonian model for the more phenomenological non-Markovian noise model we described earlier and the assumption (2.1). Furthermore, if we specify r coarse-grained locations, \mathcal{I}_r , and by letting $F(\mathcal{I}_r)$ denote the sum over all time-resolved fault paths with faults at those r coarse-grained locations (with the time-resolved fault paths unrestricted elsewhere), then we similarly obtain $\|F(\mathcal{I}_r)\|_{\text{sup}} \leq \prod_{j \in \mathcal{I}_r} \tau\|H_{SB,j}\|_{\text{sup}} \leq (\lambda_0\tau)^r$ [49]. Hence, the Hamiltonian noise model described by equations (2.2) and (2.3) and the condition (2.6) is local with strength $\varepsilon \equiv \lambda_0\tau$.

2.3 Fault-Tolerant Quantum Circuit Simulations

In the fault-tolerant implementation of some ideal quantum computation, each operation will be *simulated* by an encoded operation that acts on the logical qubits of a quantum error-correcting commute.

code. Encoded operations are implemented by simulation procedures or, simply, *gadgets* that are composite objects and consist of elementary noisy physical operations. Apart from gadgets that implement encoded operations, there also exist gadgets that implement quantum error correction. In describing the operation of gadgets, we will speak of the number of *faults* that are *contained* in them; as defined previously, we say that a fault has occurred in a location inside a gadget when the actual noisy operation at that location deviates from the ideal operation. On the other hand, the result of faults is to cause *errors* in the logical state of the code blocks that are processed by the gadgets.

Roughly speaking, gadgets are designed to prevent errors introduced by faults inside them from propagating to become many more errors within the same code block. To make this intuitive description precise, I will first define what is meant by *simulation* of some ideal operation and also what property makes a simulation *fault-tolerant*. Then, I will describe certain key properties that gadgets realizing such simulations need to satisfy. However, I will postpone the discussion about how gadgets with these properties are constructed until chapter 5.

For the discussion that follows some notation will be helpful: Elementary physical operations will be called *level-0 gadgets* or *0-Gas*. Our 0-Gas include the quantum gates of a finite quantum universal gate set, the preparation of a single qubit in the $|0\rangle$ state or *0-preparation*, and the measurement of a single qubit in the computation basis that records the classical outcome or *0-measurement*. Although this set of 0-Gas is sufficient, it will be convenient for our gadget constructions in chapter 5 to add a 0-Ga that prepares a single qubit in the $+1$ eigenstate of X , $|+\rangle$, and another 0-Ga that measures a single qubit along the eigenbasis of X and records the classical outcome.

For each 0-Ga, there is a gadget implementing the corresponding *encoded* (or *logical*) operation; such gadgets will be called *level-1 gadgets* or *1-Gas* and they act on *level-1 code blocks* or *1-blocks*. Gadgets that implement quantum error correction will be called *level-1 error-correction gadgets* or *1-ECs*. Here, the term *level* is used to indicate whether coding has been used: physical operations are not encoded, whereas gadgets use the encoding of some quantum code.⁴

To obtain the fault-tolerant simulation, we replace each 0-Ga in the ideal circuit by the corresponding 1-Ga *followed* by a 1-EC in all output 1-blocks. The exception is measurement 0-Gas that have no quantum output and which are replaced by the corresponding 1-Gas alone. We call each 1-Ga grouped together with the 1-ECs succeeding it a *level-1 rectangle* or *1-Rec* (measurement 1-Gas form 1-Recs by themselves). Figure 2.1 shows the replacement rule for a two-qubit 0-Ga.

These replacement rules generate the fault-tolerant simulation of any ideal quantum circuit. To make precise the notion of simulation, it is helpful to introduce the notion of an *ideal decoder* or *i-decoder*. This device performs *ideal* (i.e., faultless) error correction on its input 1-block and subsequently decodes the 1-block to a single qubit tracing over (i.e., discarding) all other output

⁴In the next chapter, we will also define *level-2 gadgets*, etc., and the notation anticipates this fact.

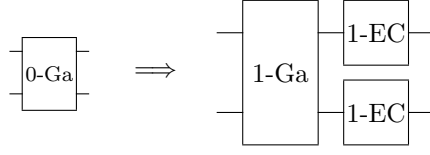


Figure 2.1. A schematic of the replacement rules generating the fault-tolerant simulation applied to a two-qubit gate. Each 0-Ga in the ideal circuit is replaced by a 1-Rec which consists of the 1-Ga corresponding to the 0-Ga followed by a 1-EC acting on each 1-block output from the 1-Ga.

ancillary qubits. It should be emphasized that i-decoders are useful as mathematical *tools* in our analysis but do *not* correspond to any physical operation applied during the simulation⁵. Let the 1-Rec that simulates some ideal 0-Ga be called *correct*; the following definition explains what is meant by such a simulation.

Definition 5 (Correctness). *A 1-Rec is correct if the 1-Rec followed by i-decoders is equivalent to the i-decoders followed by the ideal 0-Ga that the 1-Rec simulates. For a single-qubit gate 0-Ga, correctness means schematically:*

$$\boxed{\text{correct 1-Rec}} \text{---} \boxed{\text{i-decoder}} \text{---} = \text{---} \boxed{\text{i-decoder}} \text{---} \boxed{\text{ideal 0-Ga}} \text{---} .$$

In order words, suppose that an ideal 0-Ga applies the unitary transformation U to r qubits. If the input to the corresponding 1-Rec is r 1-blocks in, say, the state $|\psi\rangle_L$, applying i-decoders to each of the r 1-blocks maps $|\psi\rangle_L$ to $|\psi\rangle$. Then, if the 1-Rec is correct, the i-decoders following it map the output of the 1-Rec to $U|\psi\rangle$. In this sense, correctness captures the idea of what it means for the 1-Rec to simulate the corresponding ideal 0-Ga.

Definition 5 applies to 1-Recs simulating gate 0-Gas and requires a minor adaptation for single-qubit preparation or measurement 1-Recs. For preparation 1-Recs, correctness is meant as the property that a preparation 1-Rec followed by an i-decoder is equivalent to the ideal preparation 0-Ga—a correct preparation 1-Rec *annihilates* an i-decoder, or schematically:

$$\boxed{\text{correct prep. 1-Rec}} \text{---} \boxed{\text{i-decoder}} \text{---} = \boxed{\text{ideal 0-prep.}} \text{---} .$$

For measurement 1-Recs (which consist of the measurement 1-Gas alone), correctness is meant as the property that a measurement 1-Rec is equivalent to an i-decoder followed by the ideal measurement 0-Ga—a correct measurement 1-Ga *creates* an i-decoder, or schematically:

$$\text{---} \boxed{\text{correct meas. 1-Rec}} = \text{---} \boxed{\text{i-decoder}} \text{---} \boxed{\text{ideal 0-meas.}} .$$

⁵Although, certainly, noisy decoders can be constructed as physical devices.

Now that we have translated the idea of gate simulation to the notion of correctness, we are ready to discuss what it means for a simulation to be fault-tolerant. Intuitively, a simulation is fault-tolerant if it is robust against faults. More concretely, if we use a quantum error-correcting code that corrects arbitrary Pauli errors on up to t qubits in a 1-block, we expect that a noisy simulation will be robust against faults that act nontrivially on no more than t locations.

To make this requirement even more precise, we will need to consider *level-1 extended rectangles* or *1-exRecs* that are formed by combining each gate 1-Ga with the 1-ECs *both* preceding and following it. In the case of preparation 1-Gas, the 1-exRec is the 1-Ga grouped with the 1-ECs following it since there is no quantum input; for measurement 1-Gas, the 1-exRec is the 1-Ga grouped with the 1-ECs preceding it since there is no quantum output. According to this definition, we observe that the 1-exRecs that simulate two successive 0-Gas have at least one 1-EC in common; we will say that these 1-exRecs *overlap*. For instance, figure 2.2 shows the 1-exRecs corresponding to the fault-tolerant simulation of two successive two-qubit gate 0-Gas.

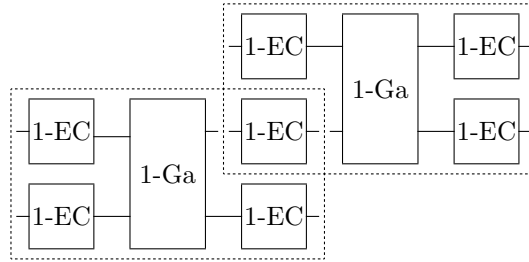


Figure 2.2. The two 1-exRecs, indicated by dashed lines, that simulate two succeeding gate 0-Gas are overlapping since they share a 1-EC which is a trailing 1-EC of the earlier 1-exRec and a leading 1-EC of the later 1-exRec.

The intuitive reason why we need to consider 1-exRecs instead of 1-Recs when considering fault tolerance is the following: Whether a 1-Rec is correct or not depends not only on the faults that it contains but also on the state of its input 1-blocks. In turn, the state of these input 1-blocks depends on the faults inside the 1-ECs preceding the 1-Rec. Although one could imagine that other faults in the past of the computation are also important, we will show that this is not the case: The correctness of a 1-Rec only depends on how many faults have occurred inside the 1-exRec that contains it. We will say that a 1-exRec which does not contain “too many” faults is *good*; the following definition explains how many is too many.

Definition 6 (Goodness). *Consider a fault-tolerant quantum circuit simulation based on a distance- d quantum error-correcting code, and let $t = \lfloor \frac{d-1}{2} \rfloor$. A 1-exRec is good if it contains at most t faults; if it is not good it is bad.*

The main result in this chapter is that the goodness of a 1-exRec implies the correctness of the 1-Rec it contains. The proof of this result relies on the existence of explicit constructions of gadgets

(i.e., 1-Gas and 1-ECs) with suitable properties. To state these properties, let us introduce the notion of an *ideal s-filter* or simply *s-filter*, where s is an integer between zero and n , the number of physical qubits in the 1-block. An s -filter is a device that performs an *ideal* (i.e., faultless) orthogonal projection onto the subspace spanned by syndromes corresponding to Pauli errors acting on at most s qubits in the 1-block; i.e., it leaves the 1-block *unchanged* if a *hypothetical* ideal syndrome measurement would *always* indicate a recovery Pauli operator of weight no more than s —it does *not* perform syndrome *measurement* and does *not* apply the recovery operator—and rejects the 1-block (and the whole computation) otherwise. If a 1-block is not rejected by an s -filter, we say that the 1-block *passes through* the s -filter. Just as i-decoders, s -filters are *tools* in our analysis and do not correspond to any physical operation applied during the fault-tolerant simulation. We will use s -filters as *tags* we can insert in our quantum circuit diagrams to indicate that the syndrome at some given point would indicate no more than s Pauli errors *assuming* faultless syndrome measurement were performed at that point; we will be careful to insert these tags so that they act as the identity on the actual computation.

For brevity, let a 1-Ga or 1-EC that contains no more than r faults be called *r-good*. The key gadget properties starting with the properties for 1-EC gadgets are as follows.

Properties 1 (1-EC Properties). *Let $t = \lfloor \frac{d-1}{2} \rfloor$. 1-EC gadgets for a distance- d quantum error-correcting code satisfy:*

(a) $\forall r \leq t$, the output of an r -good 1-EC passes through an r -filter, or

$$\boxed{\text{\textit{r-good 1-EC}}} = \boxed{\text{\textit{r-good 1-EC}}} \boxed{\text{\textit{r-filter}}} ,$$

(b) $\forall s, r$ such that $s + r \leq t$, if a 1-block which passes through an s -filter is input to an r -good 1-EC followed by an i-decoder, the output is the same as with the r -good 1-EC omitted, or

$$\boxed{\text{\textit{s-filter}}} \boxed{\text{\textit{r-good 1-EC}}} \boxed{\text{\textit{i-decoder}}} = \boxed{\text{\textit{s-filter}}} \boxed{\text{\textit{i-decoder}}} .$$

Property 1(a) is the statement that on an *arbitrary* input, a 1-EC which contains at most r faults produces an output on which ideal syndrome measurement would find at most r Pauli errors. Property 1(b) states that if the sum of the Pauli errors in the input 1-block and the faults in the 1-EC is no greater than t , then ideal decoding of the output gives the same output as if the 1-EC were not present (or, since the i-decoder performs ideal error correction itself, if the 1-EC were present but contained no faults). Both properties reflect the requirement that $r \leq t$ faults inside a 1-EC gadget must not lead to errors which propagate to more than r qubits in the 1-block.

The properties for 1-Ga gadgets are:

Properties 2 (1-Ga Properties). *Let $t = \lfloor \frac{d-1}{2} \rfloor$. Consider k 1-blocks which respectively pass through $\{s_1, s_2, \dots, s_k\}$ -filters and let them be input to an r -good 1-Ga gadget for a distance- d quantum error-correcting code simulating a k -qubit gate, where $s \equiv \sum_{i=1}^k s_i + r \leq t$. Then,*

(a) *all output 1-blocks pass through s -filters, or*

$$\boxed{\{s_i\}\text{-filters}} \text{---} \boxed{r\text{-good 1-Ga}} \text{---} = \boxed{\{s_i\}\text{-filters}} \text{---} \boxed{r\text{-good 1-Ga}} \text{---} \boxed{s\text{-filters}} \text{---} ,$$

(b) *the r -good 1-Ga followed by i -decoders is equivalent to applying the i -decoders first followed by the ideal k -qubit 0-Ga, or*

$$\boxed{\{s_i\}\text{-filters}} \text{---} \boxed{r\text{-good 1-Ga}} \text{---} \boxed{i\text{-decoders}} \text{---} = \boxed{\{s_i\}\text{-filters}} \text{---} \boxed{i\text{-decoders}} \text{---} \boxed{\text{ideal 0-Ga}} \text{---} .$$

For the properties above, the s_i -filter is applied to the i th input 1-block and, for property 2(a), each one of the k s -filters on the right is applied to one of the k output 1-blocks.

Property 2 requires a minor adaptation for preparation and measurement 1-Gas. For 1-preparations and $\forall r \leq t$, property 2(a) becomes:

$$\boxed{r\text{-good 1-prep.}} \text{---} = \boxed{r\text{-good 1-prep.}} \text{---} \boxed{r\text{-filter}} \text{---} ,$$

and property 2(b) becomes:

$$\boxed{r\text{-good 1-prep.}} \text{---} \boxed{i\text{-decoder}} \text{---} = \boxed{\text{ideal 0-prep.}} \text{---} .$$

For 1-measurements, which have classical bits as outputs, there is no analog of property 2(a) and, $\forall r, s$ such that $r + s \leq t$, property 2(b) becomes:

$$\boxed{s\text{-filter}} \text{---} \boxed{r\text{-good 1-meas.}} = \boxed{s\text{-filter}} \text{---} \boxed{i\text{-decoder}} \text{---} \boxed{\text{ideal 0-meas.}} ,$$

where the classical processing of the measurement outcomes is either performed in encoded form using a classical error-correcting code or, if the outcomes are decoded to the level of single classical bits, the classical gates that perform the decoding are faultless. Otherwise, a single fault in the final classical decoding step (e.g., a fault in our *reading* of the final outcome from the display of a detector) could cause an error in the outcome and the property is not satisfied.

It is now time to state and prove the basic lemma of this chapter.

Lemma 1 (Goodness Implies Correctness). *Consider a fault-tolerant quantum circuit simulation based on a distance- d quantum error-correcting code that is executed using gadgets satisfying properties 1 and 2. Then, the 1-Rec contained in a good 1-exRec is correct. For a good 1-exRec corre-*

sponding to a single-qubit gate 0-Ga, the lemma means schematically:

$$\text{---} \boxed{\text{1-EC}} \text{---} \boxed{\text{1-Ga}} \text{---} \boxed{\text{1-EC}} \text{---} \boxed{\text{i-decoder}} \text{---} = \text{---} \boxed{\text{1-EC}} \text{---} \boxed{\text{i-decoder}} \text{---} \boxed{\text{ideal 0-Ga}} \text{---} .$$

Proof. We will only give the proof for single-qubit gate 0-Gas—the argument is essentially the same for multi-qubit gate, preparation or measurement 0-Gas. By definition, a 1-exRec is good if it contains no more than $t = \lfloor \frac{d-1}{2} \rfloor$ faults. Suppose there are s faults in the leading 1-EC, r faults in the 1-Ga, and s' faults in the trailing 1-EC, with $s + r + s' \leq t$. Then, we can write the left-hand side of our equation schematically as:

$$\text{---} \boxed{\begin{smallmatrix} s\text{-good} \\ \text{1-EC} \end{smallmatrix}} \text{---} \boxed{\begin{smallmatrix} r\text{-good} \\ \text{1-Ga} \end{smallmatrix}} \text{---} \boxed{\begin{smallmatrix} s'\text{-good} \\ \text{1-EC} \end{smallmatrix}} \text{---} \boxed{\text{i-decoder}} \text{---} .$$

Using property 1(a), we can insert an s -filter after the leading 1-EC and then, using property 2(a), we can insert an $(s+r)$ -filter after the 1-Ga to obtain the equivalent circuit:

$$= \text{---} \boxed{\begin{smallmatrix} s\text{-good} \\ \text{1-EC} \end{smallmatrix}} \text{---} \boxed{s\text{-filter}} \text{---} \boxed{\begin{smallmatrix} r\text{-good} \\ \text{1-Ga} \end{smallmatrix}} \text{---} \boxed{\begin{smallmatrix} (s+r)\text{-} \\ \text{filter} \end{smallmatrix}} \text{---} \boxed{\begin{smallmatrix} s'\text{-good} \\ \text{1-EC} \end{smallmatrix}} \text{---} \boxed{\text{i-decoder}} \text{---} .$$

Using property 1(b), we can now omit the trailing 1-EC and then, using property 2(a) backwards, we can omit the $(s+r)$ -filter that follows the 1-Ga to obtain the equivalent circuit:

$$= \text{---} \boxed{\begin{smallmatrix} s\text{-good} \\ \text{1-EC} \end{smallmatrix}} \text{---} \boxed{s\text{-filter}} \text{---} \boxed{\begin{smallmatrix} r\text{-good} \\ \text{1-Ga} \end{smallmatrix}} \text{---} \boxed{\text{i-decoder}} \text{---} .$$

Finally, using property 2(b), we can move the i-decoder to the left thereby converting the 1-Ga to the ideal 0-Ga it simulates and then, using property 1(a) backwards, we can remove the s -filter that follows the leading 1-EC. We thus obtain the equivalent circuit:

$$= \text{---} \boxed{\begin{smallmatrix} s\text{-good} \\ \text{1-EC} \end{smallmatrix}} \text{---} \boxed{\text{i-decoder}} \text{---} \boxed{\text{ideal 0-Ga}} \text{---} .$$

This proves the lemma. □

2.4 Accuracy

Lemma 1 implies that if all 1-exRecs are good, the fault-tolerant simulation will produce exactly the same outcome probability distribution as the ideal simulated circuit: If all measurement 1-exRecs are good, then we can replace the 1-Recs contained in them by equivalent circuits of i-decoders followed by the ideal simulated measurement 0-Gas. Next, if all gate 1-exRecs are good, we can propagate these i-decoders to the left thereby transforming all gate 1-Recs to equivalent circuits applying

the ideal simulated gate 0-Gas. Finally, if all preparation 1-exRecs are good, we can annihilate the i-decoders thereby obtaining an equivalent circuit with ideal preparation 0-Gas. Overall, if all 1-exRecs are good, the fault-tolerant simulation is equivalent to the ideal simulated circuit. Put simply, what happens is that the goodness of all 1-exRecs constrains the faults inside gadgets to be such that they never cause errors on more than t qubits in any 1-block; hence, errors are always successfully corrected and the fault-tolerant simulation as a whole is successful.

Stochastic Local Noise

Suppose that we simulate an ideal circuit of size L , and suppose noise is local and stochastic with strength p . For a fault-tolerant simulation based on a distance- d error-correcting code to be unsuccessful, at least $t + 1$ faults must have occurred inside one of the L 1-exRecs where $t = \lfloor \frac{d-1}{2} \rfloor$. Since the probability of all fault paths with at least $t + 1$ faults in *specified* locations is at most p^{t+1} , if we let C denote the number of locations in the largest 1-exRec, the probability that a specified 1-exRec contains at least $t + 1$ faults is at most $\binom{C}{t+1} p^{t+1}$. This implies that the error, δ , of the fault-tolerant simulation can be upper bounded as in equation (1.15) by substituting $p \rightarrow \binom{C}{t+1} p^{t+1}$,

$$\delta \leq 2L \binom{C}{t+1} p^{t+1}. \quad (2.7)$$

Hence, the fault-tolerant simulation has an improved accuracy compared to the unencoded computation provided $\binom{C}{t+1} p^{t+1} < p$, or $p < p_{\text{crit}}$ where

$$p_{\text{crit}} \equiv \left(\binom{C}{t+1} \right)^{-1/t}. \quad (2.8)$$

Local Noise

We can arrive at a similar conclusion if we consider local noise that is not stochastic. In this case, we consider the fault-tolerant simulation as a noisy quantum circuit represented by a unitary acting between the system qubits and a common bath. By performing a Lie-Trotter expansion of the Hamiltonian, H_{SB} , that couples the system and the bath, we can express the evolution at some specific coarse-grained location and the bath as a sum of an operator that acts as the identity on the coarse-grained location *plus* a fault operator acting nontrivially on the coarse-grained location. Repeating for all coarse-grained locations, we obtain a fault-path expansion of the entire fault-tolerant simulation.

We can now group fault paths in two terms: The first term, $Gd^{(1)}$, is a sum over all *good* fault paths; a fault path is good if it applies a fault operator in no more than t coarse-grained locations within each of the L 1-exRecs. Lemma 1 implies that all good fault paths will give identical outcome statistics as the ideal computation; hence, their sum $Gd^{(1)}$ will also give the ideal outcome statistics.

The second term, $Bd^{(1)}$, is a sum over all remaining fault paths which we will call *bad*. A fault path is bad if it applies a fault in at least $t + 1$ coarse-grained locations inside at least one of the L 1-exRecs. The two terms $Gd^{(1)}$ and $Bd^{(1)}$ play an analogous role to the Gd and Bd operators in §1.2.4; specifically, the error, δ , of the fault-tolerant simulation can be upper bounded by the sup norm of the sum over all bad fault paths, $\delta \leq 2\|Bd^{(1)}\|_{\text{sup}}$.

In the special case when local noise is due to a noise process that satisfies condition (2.1), upper bounding $\|Bd^{(1)}\|_{\text{sup}}$ is straightforward: We order the L 1-exRecs in some arbitrary way and we also consider some ordering of the coarse-grained locations inside each 1-exRec. We now organize $Bd^{(1)}$ as a sum of L terms where for the i th term the i th 1-exRec is bad, all 1-exRecs preceding it are good and we put no restriction on the 1-exRecs following it. Then, we express each of these L terms as a sum of terms where for each term a different set, \mathcal{I}_{t+1} , of $t + 1$ faulty coarse-grained locations is identified inside the bad 1-exRec, the ideal gates are applied in all coarse-grained locations that precede the latest of the $t + 1$ faults and are not in \mathcal{I}_{t+1} and the *unitary* coupling described by H_{SB} is applied between the bath and all coarse-grained locations that are not directly coupled by H_{SB} to any of the $t + 1$ faults and are subsequent to the latest of them. There are at most $\binom{C}{t+1}$ terms of this form and each has sup norm at most ε^{t+1} . Thus, we conclude that

$$\|Bd^{(1)}\|_{\text{sup}} \leq L \binom{C}{t+1} \varepsilon^{t+1}. \quad (2.9)$$

Therefore, the error of the fault-tolerant simulation is smaller than the error, $2L\varepsilon$, of the unencoded computation provided $\varepsilon < \varepsilon_{\text{crit}}$ where

$$\varepsilon_{\text{crit}} \equiv \left(\binom{C}{t+1} \right)^{-1/t}. \quad (2.10)$$

To obtain an upper bound on $Bd^{(1)}$ for general local noise is more challenging. The essential difference is that now condition (2.1) does not necessarily hold for every coarse-grained location irrespective of whether other coarse-grained locations are faulty or ideal. Instead, definition 4 only allows us to upper bound the sup norm of the *sum* of all fault paths with faults at some r specific coarse-grained locations⁶. In particular, the method we used in the previous paragraph to organize $Bd^{(1)}$ as a sum of at most $L \binom{C}{t+1}$ terms does not immediately lead to the desired bound: The problem is that in each such term we have imposed a restriction on some locations outside the set \mathcal{I}_{t+1} ; namely that the *ideal* gates are applied in all coarse-grained locations that precede the latest of the $t + 1$ faults. Hence, each such term is *not* a sum over *all* fault paths with faults at the locations in \mathcal{I}_{t+1} (i.e., some fault paths are excluded) and, so, we cannot directly upper bound its sup norm by ε^{t+1} .

⁶Note that this summation corresponds to considering the unitary coupling described by H_{SB} between the bath and all other coarse-grained locations not coupled directly with the r specified faulty ones.

The solution to this problem is based on the idea that we can perform our Lie-Trotter expansion “in reverse,” i.e., we can reexpress the ideal realization of a given coarse-grained location as the unitary coupling described by H_{SB} *minus* a fault operator that acts on this coarse-grained location. Our motivation for doing this is that in this way we can eliminate the restriction that we imposed above on some locations outside the set \mathcal{I}_{t+1} to be ideal. In the following lemma we develop this idea and perform the combinatorics.

Lemma 2 (Fault Path Inclusion-Exclusion). *Consider a noisy quantum circuit with C locations which is subject to local noise with strength ε . We express the unitary transformation acting on the system and the bath as a sum of fault paths where, for each fault path, each location is either faulty or ideal, and let $F(\mathcal{I}_{r \geq s})$ denote the sum over all fault paths with at least s faults. Then,*

$$\|F(\mathcal{I}_{r \geq s})\|_{\text{sup}} \leq \binom{C}{s} \varepsilon^s e^{(C-s)\varepsilon}. \quad (2.11)$$

Proof. Let us choose some arbitrary ordering of the C coarse-grained locations that does not need to correspond to the actual time ordering of the operations in the quantum circuit—speaking of *earlier* or *later* locations should be understood as referring to this ordering. Then, $F(\mathcal{I}_{r \geq s})$ can be expressed as a sum of at most $\binom{C}{s}$ terms, where each term corresponds to a sum over all fault paths such that a particular set, \mathcal{I}_s , of s faulty coarse-grained locations is identified and the ideal gates are applied in all coarse-grained locations that precede the latest of the s specified faults and are not in \mathcal{I}_s (i.e., the different fault paths in the sum act identically in the s specified locations and the locations preceding them and may differ on the coarse-grained locations that succeed the latest of the s specified faults). Let us say that this is the *original expansion* of $F(\mathcal{I}_{r \geq s})$.

Next, in each term of the original expansion, we consider the ideal gates that have been applied to locations preceding the latest of the s faults. We proceed to reexpress all of them as the unitary transformation that couples those locations and the bath *minus* a fault operator. We thus obtain the *derived expansion* which can be expressed as a sum of terms, where each term corresponds to a sum over all fault paths such that m specific coarse-grained locations are identified as faulty with $s \leq m \leq C$ and without other restrictions. We now need to count how many times a fault path where m faulty coarse-grained locations have been specified appears in this derived expansion.

Let $F(\mathcal{I}_{r=m})$ denote the sum over all the $\binom{C}{m}$ different terms in the derived expansion that identify exactly m faulty coarse-grained locations. Each of these terms may arise from at most $\binom{m-1}{s-1}$ terms in the original expansion, since we need to count the number of ways of choosing which $s-1$ of the first $m-1$ faulty coarse-grained locations belonged in the original expansion and the maximum, $\binom{m-1}{s-1}$, corresponds to the case when each faulty coarse-grained location is coupled independently with the bath. Therefore, the sup norm of each of these terms is at most $\binom{m-1}{s-1} \varepsilon^m$, and so the sup norm of $F(\mathcal{I}_{r=m})$ can be upper bounded by $\binom{C}{m} \binom{m-1}{s-1} \varepsilon^m$ (the sign of each term in

the derived expansion may be $+1$ or -1 , but these signs have no consequence for our bound).

Since our derived expansion expresses $F(\mathcal{I}_{r \geq s})$ as $\sum_{m=s}^C F(\mathcal{I}_{r=m})$, we conclude that

$$\begin{aligned} \|F(\mathcal{I}_{r \geq s})\|_{\text{sup}} &\leq \sum_{m=s}^C \binom{C}{m} \binom{m-1}{s-1} \varepsilon^m = \binom{C}{s} \varepsilon^s \sum_{m=s}^C \frac{s}{m} \binom{C-s}{m-s} \varepsilon^{m-s} \\ &\leq \binom{C}{s} \varepsilon^s \sum_{m=0}^{\infty} \frac{(C-s)^m}{m!} \varepsilon^m = \binom{C}{s} \varepsilon^s e^{(C-s)\varepsilon}. \end{aligned} \quad (2.12)$$

This proves the lemma. □

We can now use lemma 2 twice to upper bound $\|Bd^{(1)}\|_{\text{sup}}$. Recall that $Bd^{(1)}$ is the sum of all fault paths with at least $t+1$ faults in at least one of the L 1-exRecs. By choosing an ordering of the L 1-exRecs, our first step is the same as before: we write $Bd^{(1)}$ as a sum of terms where for the i th term the i th 1-exRec is bad, all 1-exRecs preceding it are good and we put no restriction on the succeeding 1-exRecs; this is now our *original expansion*. We next reexpress every good 1-exRec as a sum of one term where we apply the unitary coupling between the bath and the coarse-grained locations inside the 1-exRec *minus* a second term which contains all fault paths that make the 1-exRec to be *bad*; this leads to our *derived expansion*. By using the same argument as in the proof of lemma 2, we can express $Bd^{(1)}$ as a sum of terms where for each term m specific 1-exRecs are bad with $1 \leq m \leq L$ and we put no restriction on the other 1-exRecs. It follows that we can upper bound the sup norm of $Bd^{(1)}$ as in equation (2.11) with $C \rightarrow L$, $s \rightarrow 1$, and $\varepsilon \rightarrow \|F(\mathcal{I}_{r \geq s})\|_{\text{sup}}$;

$$\|Bd^{(1)}\|_{\text{sup}} \leq L \|F(\mathcal{I}_{r \geq t+1})\|_{\text{sup}} \cdot e^{(L-1) \cdot \|F(\mathcal{I}_{r \geq t+1})\|_{\text{sup}}}, \quad (2.13)$$

where $F(\mathcal{I}_{r \geq t+1})$ denotes the sum of all fault paths that make a specific 1-exRec bad since all these fault paths must contain at least $t+1$ faults. Assuming ε is sufficiently small so that $(L-1) \cdot \|F(\mathcal{I}_{r \geq t+1})\|_{\text{sup}} \leq 1$ and by using lemma 2 again to upper bound $\|F(\mathcal{I}_{r \geq t+1})\|_{\text{sup}}$,

$$\|Bd^{(1)}\|_{\text{sup}} \leq eL\kappa \binom{C}{t+1} \varepsilon^{t+1}, \quad (2.14)$$

where $\kappa \geq e^{(C-t-1)\varepsilon}$. The error, δ , of the fault-tolerant simulation can then be upper bounded by

$$\delta \leq 2\|Bd^{(1)}\|_{\text{sup}} \leq 2eL\kappa \binom{C}{t+1} \varepsilon^{t+1}, \quad (2.15)$$

which is smaller than the error, $2L\varepsilon \cdot e^{(L-1)\varepsilon} \leq 2eL\varepsilon$ assuming $(L-1)\varepsilon \leq 1$, of the unencoded computation provided $\kappa \binom{C}{t+1} \varepsilon^{t+1} < \varepsilon$, or $\varepsilon < \varepsilon_{\text{crit}}$ where

$$\varepsilon_{\text{crit}} \equiv \left(\kappa \binom{C}{t+1} \right)^{-1/t}, \quad (2.16)$$

for any $\kappa \geq e^{(C-t-1)\varepsilon_{\text{crit}}}$.

In obtaining the derived expansion of $Bd^{(1)}$ above, there is a detail that we have ignored: since two successive bad 1-exRecs overlap, their badness can be due to sets of faults that are not disjoint. Although this seems to complicate the analysis, we can in fact modify our definition of goodness in such a way that any two successive bad 1-exRecs will be designated to be simultaneously bad only due to sets of faults that are disjoint. With this modified definition which we will give in §3.2, our calculation above is valid.

2.5 History and Acknowledgements

The basic principles of fault-tolerant quantum circuit simulations and explicit gadget constructions satisfying properties 1 and 2 (to be discussed in chapter 5) were first given by Shor [51]. In this and in all subsequent work, the properties of fault-tolerant gadgets were discussed *semantically*, i.e., in terms of an explicit representation of the errors afflicting the quantum state being processed by these gadgets.

After Gottesman, Preskill and I began thinking about the problem of proving the quantum threshold theorem for distance-3 codes, it quickly became clear that the criterion for fault tolerance would involve extended rectangles whose analysis we found to be difficult semantically. In fact, our definition of goodness had previously been used by Knill, Laflamme and Zurek [52], but their argument was informal and it was not at all clear how to apply it to recursive fault-tolerant simulations (to be discussed in chapter 3). During and shortly after the QIP 2005 workshop, Gottesman suggested that we use an insight from [53] to phrase our definition of correctness in terms of ideal decoding circuits; this *syntactic* definition led to the formulation of lemma 1.

An analysis of fault-tolerant quantum computation for a coherent noise model was first formulated in an insightful paper by Terhal and Burkard [49]. Their analysis considered a noise model where a locality condition is imposed not only on the interaction between the system and the bath, but also on interactions among the degrees of freedom within the bath. The essential ingredient for generalizing the proof in [49] was the formulation of lemma 2 for the Hamiltonian noise model described in §2.2. Lemma 2 for this noise model was proved by Preskill and it was included in [1] with a simplified proof due to Gottesman. The version of this lemma given here applies to general local noise; the proof is essentially the same as in [1].

Chapter 3

Recursive Simulations and the Quantum Threshold Theorem

3.1 Introduction

In this chapter, I will discuss recursive fault-tolerant quantum circuit simulations and prove the quantum threshold theorem for local noise. The idea of recursive simulations is intuitive: In the previous chapter, I described how any ideal quantum circuit can be simulated using gadgets that perform the computation in an encoded form using some quantum error-correcting code. Furthermore, these simulations are fault-tolerant in the sense that the noisy simulation is more accurate than the noisy unencoded computation provided the strength of local noise is below some critical value. But, if for some noise strength the fault-tolerant simulation improves the accuracy of the unencoded computation, we expect that a fault-tolerant simulation of the fault-tolerant simulation will improve it even more! And we can continue applying this idea again and again building a hierarchy of simulations that achieves any desired accuracy. We will then identify the critical noise strength as the *accuracy threshold* for quantum computation.

By construction, at any level, k , of a recursive fault-tolerant simulation, the simulation will be mapped to the corresponding simulation at level $k-1$ if we replace every 1-Rec inside it by a 0-Ga. If we replace every 1-Rec inside a simulation at level $k-1$ by a 0-Ga, the simulation at level $k-1$ will be mapped to the corresponding simulation at level $k-2$, etc. Our analysis of the accuracy of recursive fault-tolerant simulations will exploit exactly this property: We will show that the fault-tolerant simulation at level k with noise strength ε is equivalent to another fault-tolerant simulation at level $k-1$ with a *transformed* noise strength, $\varepsilon^{(1)}$. Then, by applying the same argument a second time, the fault-tolerant simulation at level $k-1$ with noise strength $\varepsilon^{(1)}$ is equivalent to another fault-tolerant simulation at level $k-2$ with a *transformed* noise strength, $\varepsilon^{(2)}$. After we apply this argument k times, we will obtain an equivalent simulation at level 0 with noise strength $\varepsilon^{(k)}$; but, a level-0 simulation is an execution of the ideal circuit with noisy physical operations! We conclude

that the initial fault-tolerant simulation at level k is equivalent to an execution of the ideal quantum circuit with operations which have noise strength $\varepsilon^{(k)}$. Provided the physical noise strength, ε , is below the accuracy threshold value, $\varepsilon^{(k)}$ can be made arbitrarily small, thereby showing that the level- k simulation can achieve any desired accuracy. This is the outline of our proof of the quantum threshold theorem.

The remaining of this chapter is organized as follows: In §3.2, I discuss the replacement rules that give rise to recursive fault-tolerant simulations. Then, in §3.3, I prove the *level-reduction* lemma which shows that recursive fault-tolerant simulations subject to local noise can be analysed “one level at a time.” Finally, in §3.4, I use this lemma to prove the quantum threshold theorem for local noise. I also discuss a method for obtaining improved rigorous lower bounds on the quantum accuracy threshold by means of a more detailed combinatorial analysis; such an analysis is performed explicitly in chapter 5 for a particular code.

3.2 Recursive Fault-Tolerant Simulations

The quantum computation performed by a recursive fault-tolerant simulation is protected by a *concatenated* quantum code [54]. The code block of a concatenated code is constructed as a hierarchy of codes within codes—the code block at any level, k , of this hierarchy is built from logical qubits encoded at level $k - 1$ of the hierarchy. And there exists a corresponding hierarchy of gadgets: At the physical level, which we may call *level 0*, gadgets are just the noisy physical operations that we have called 0-Gas. At the next level of this recursion, *level 1*, each ideal gate is simulated by a 1-Rec that acts on 1-blocks of some quantum error-correcting code, \mathcal{C} , as was discussed in the previous chapter. One level higher in the recursion, *level 2*, an ideal gate is simulated by a 2-Rec that acts on the code blocks of the concatenated quantum code $\mathcal{C} \circ \mathcal{C}$; we will call such code blocks *2-blocks*. A 2-Rec is constructed by replacing each 0-Ga inside a 1-Rec by the 1-Rec corresponding to the 0-Ga. At level k of the recursion, ideal gates are simulated by k -Recs that are constructed by replacing each 0-Ga inside a $(k-1)$ -Rec by the corresponding 1-Rec; k -Recs act on k -blocks, i.e., code blocks of a quantum code that is concatenated k times. If k levels of recursion are used in total, then there will be one k -Rec for each location in the ideal quantum circuit we simulate. We will often call the computation realized by k -Recs a *level- k simulation*; in particular, a level-0 simulation is a direct implementation of some quantum circuit with noisy 0-Gas. Figure 3.1 shows a schematic of a quantum circuit executing a recursive simulation.

Lemma 1 in the previous chapter established that the 0-Recs inside good 1-exRecs (i.e., 1-exRecs that contain no more than $t = \lfloor \frac{d-1}{2} \rfloor$ faults, where d is the distance of the quantum code used) are correct (i.e., they perform an accurate simulation of the corresponding 0-Ga). This allowed us to conclude that if all 1-exRecs are good, the level-1 simulation realized by 1-Recs can be replaced by

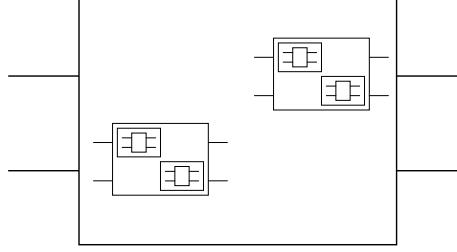


Figure 3.1. A recursive simulation. A level- k gadget is built from level- $(k-1)$ gadgets, which are built from level- $(k-2)$ gadgets, etc.

an equivalent level-0 simulation where all 1-Recs are converted to the ideal 0-Gas they simulate.

Now, we would like to generalize this conclusion. We would like to prove that no matter how many faults have occurred, the level-1 simulation can always be replaced by an equivalent level-0 simulation where the 1-Rec inside every good 1-exRec has been converted to the ideal simulated 0-Ga, and the 1-Rec inside every bad 1-exRec has been converted to some faulty implementation of the simulated 0-Ga. But, from the onset, we encounter a problem in identifying bad 1-exRecs with faulty 0-Gas: Two successive 1-exRecs overlap (i.e., they contain at least one 1-EC gadget in common) and, for this reason, they can both be bad due to sets of faults that are *not* disjoint.

Suppose, e.g., that we use a distance-3 code and that two successive 1-exRecs are both bad due to one fault in the shared 1-EC, one fault in the 1-Ga of the earlier 1-exRec and one fault in the trailing 1-EC of the later 1-exRec (Fig. 3.2). These two overlapping 1-exRecs simulate two successive 0-Gas. Ordinarily, we would think of the earlier of the two 1-Recs as the gadget that simulates the earlier of the two 0-Gas. But, in this case, we may instead *remove* the shared 1-EC from the earlier 1-Rec and regard the 1-Ga contained in the earlier 1-Rec as the gadget that simulates the earlier 0-Ga; we will say that the earlier 1-exRec has been *truncated*. Because the truncated earlier 1-exRec contains only one fault, we expect that it simulates the earlier 0-Ga accurately. The shared 1-EC which we removed from the earlier 1-exRec will be added to the later 1-Rec; so, we will regard the entire later 1-exRec—rather than the 1-Rec it contains—as the gadget that simulates the later 0-Ga; and it is only this later simulation that is inaccurate. Thus, our intuition is that the two overlapping bad 1-exRecs in figure 3.2 are really no worse than a single bad 1-exRec.

More generally, we wish to regard the badness of different 1-exRecs as arising due to disjoint sets of faults. We will say that two bad 1-exRecs are *independent* if they are bad due to a disjoint set of faults; when this is not the case, we will say that they are *non-independent*. As indicated by the example above, our criterion for when two overlapping bad 1-exRecs are independent is that the earlier bad 1-exRec must *remain bad* when all shared 1-EC gadgets are removed. To reflect this requirement, we revise definition 6 as follows.

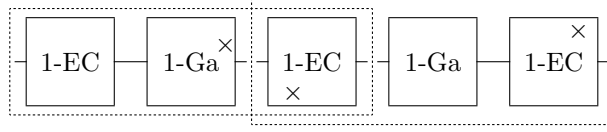


Figure 3.2. Two non-independent successive bad 1-exRecs, indicated by dashed lines, with fault locations indicated by \times . Because one of the three faults is contained in the shared 1-EC, the bad 1-exRecs are not independent events. In this situation, we may regard the earlier 1-Ga as a gadget that simulates the corresponding ideal 0-Ga accurately, and the later 1-exRec rather than the later 1-Rec as a gadget that simulates the corresponding ideal 0-Ga inaccurately.

Definition 7 (Goodness (revised)). *Consider a fault-tolerant quantum circuit simulation based on a distance- d quantum error-correcting code, and let $t = \lfloor \frac{d-1}{2} \rfloor$. A 1-exRec is good if it contains at most t faults; if it is not good it is bad. Two bad 1-exRecs are independent if they are nonoverlapping or if they overlap and the earlier 1-exRec is still bad when the shared 1-ECs are removed.*

The fact that two non-independent bad 1-exRecs can be viewed as simulating just one rather than two faulty 0-Gas will be proven rigorously in §3.3.1. But before giving more details about treating non-independent bad 1-exRecs, let us first use this fact to prove the basic level-reduction lemma.

3.3 The Level-Reduction Lemma

Our goal is to assess the error, δ , of the level- k simulation as a function of the level k of recursion. We will imagine that local noise of strength ε is acting on all elementary physical operations or, simply, *locations* realizing this level- k simulation; then the question is how δ is related to ε and k . We will answer this question indirectly: We will first show that the level- k simulation is equivalent to a level- $(k-1)$ simulation that is subject to local noise of *transformed* strength, $\varepsilon^{(1)}$. The following lemma makes this statement precise.

Lemma 3 (Level Reduction). *Consider a recursive fault-tolerant simulation that is generated by the self-similar replacement rules described above on gadgets satisfying properties 1 and 2 for a distance- d quantum error-correcting code, and let $t = \lfloor \frac{d-1}{2} \rfloor$ and C be the number of locations in the largest 1-exRec. Let the level- k simulation ($k \geq 1$) be subject to local noise of strength ε (or p if noise is stochastic). Then, there exists a level- $(k-1)$ simulation that produces the same output probability distribution and is subject to local noise of strength*

$$\varepsilon^{(1)} \leq \kappa \binom{C}{t+1} \varepsilon^{t+1},$$

where $\kappa \geq e^{(C-t-1)\varepsilon}$ (or $p^{(1)} \leq \binom{C}{t+1} p^{t+1}$ if noise is stochastic).

Proof. We start by considering all measurement 1-exRecs. Using lemma 1, we replace the 1-Rec inside every good measurement 1-exRec by an i-decoder followed by the ideal simulated measurement 0-Ga. We then continue by moving the i-decoders to the left of the 1-Rec inside every good gate 1-exRec, thereby replacing the gate 1-Recs by the ideal simulated gate 0-Gas. Finally, we annihilate all i-decoders inside good preparation 1-exRecs, thereby replacing the preparation 1-Recs by the ideal simulated preparation 0-Gas. In addition, every time i-decoders are stuck to the right of a bad gate 1-exRec, the i-decoders can be moved to the left of the *entire* bad 1-exRec which is replaced by some *faulty* implementation of the simulated gate 0-Ga; and similarly for bad measurement or preparation 1-exRecs (see §3.3.1). Overall, by means of this “wave” of i-decoders moving from the right to the left of all level-1 gadgets, we convert every 1-Rec to either an ideal or a faulty implementation of the 0-Ga it simulates. Hence, one level of coding has been removed and the initial level- k simulation has been converted to a level- $(k-1)$ simulation that produces exactly the same output.

It remains to consider the *effective* noise afflicting this level- $(k-1)$ simulation. Consider a set, $\mathcal{I}_r^{(1)}$, of r locations in the level- $(k-1)$ simulation and let $F(\mathcal{I}_r^{(1)})$ denote the sum of all fault paths with faults at all locations in $\mathcal{I}_r^{(1)}$. Each location in the level- $(k-1)$ simulation arises from a 1-Rec in the initial level- k simulation and, furthermore, r specific locations are faulty only if the corresponding 1-exRecs in the initial level- k simulation were all *independently* bad. Recall that each independent bad 1-exRec contains at least $t+1$ faults. For local stochastic noise, the sum of all fault paths with $t+1$ faults at specified locations in each of the r bad 1-exRecs has probability at most $(p^{t+1})^r$ and, so, the probability of $F(\mathcal{I}_r^{(1)})$ is at most $(p^{(1)})^r$ where

$$p^{(1)} \leq \binom{C}{t+1} p^{t+1}. \quad (3.1)$$

Hence, we conclude that the equivalent level- $(k-1)$ simulation that results after level reduction is afflicted by local stochastic noise of strength $p^{(1)}$.

For local noise that is not stochastic, we can follow the argument in proving lemma 2 to obtain the derived expansion of $F(\mathcal{I}_r^{(1)})$. Let $F_i(\mathcal{I}_{r=m_i})$ denote the sum of all fault paths such that the i th of the r specified bad 1-exRecs contains m_i faults with $t+1 \leq m_i \leq C$ (see §3.3.2 for dealing with overlapping bad 1-exRecs), and also let $m = \sum_i m_i$. Then, $F(\mathcal{I}_r^{(1)})$ can be expanded as

$$F(\mathcal{I}_r^{(1)}) = \sum_{m_1=t+1}^C \cdots \sum_{m_r=t+1}^C F_1(\mathcal{I}_{r=m_1}) \cdots F_r(\mathcal{I}_{r=m_r}), \quad (3.2)$$

so that, by performing a similar combinatorial analysis as in the proof of lemma 2,

$$\|F(\mathcal{I}_r^{(1)})\|_{\sup} \leq \sum_{m_1=t+1}^C \binom{m_1-1}{t} \binom{C}{m_1} \cdots \sum_{m_r=t+1}^C \binom{m_r-1}{t} \binom{C}{m_r} \varepsilon^m. \quad (3.3)$$

Similar to equation (2.12), we conclude that $\|F(\mathcal{I}_r^{(1)})\|_{\text{sup}} \leq (\varepsilon^{(1)})^r$ where

$$\varepsilon^{(1)} \leq \binom{C}{t+1} \varepsilon^{t+1} e^{(C-t-1)\varepsilon} . \quad (3.4)$$

Therefore, the level- $(k-1)$ simulation that results after level reduction is afflicted by local noise of strength $\varepsilon^{(1)}$, as desired.

This proves the lemma. □

3.3.1 Independent Bad Rectangles as Simulated Faults

If all 1-exRecs are good, then we can apply lemma 1 in succession to convert all of them to the ideal 0-Gas they simulate. But in the proof of the level-reduction lemma, we also need to consider the case of bad 1-exRecs and, in this case, lemma 1 does not apply. What we would like to achieve is an equivalence allowing us to convert the 1-Rec inside a bad 1-exRec that is followed by i-decoders to a circuit where i-decoders appear first and they are followed by some faulty implementation of the corresponding simulated 0-Ga. However, this is in general not possible: the 1-Rec inside a bad 1-exRec is not simulating a well-defined faulty 0-Ga independent of the input to the 1-Rec. Indeed, if we attempt to move the i-decoders from the right of the 1-Rec inside a bad 1-exRec to in front of it and thereby transform the 1-Rec to a faulty 0-Ga, the particular faulty 0-Ga we obtain may depend on the syndrome that the i-decoders find. For instance, faults in the 1-Rec might combine differently with a fault which has previously acted on the first qubit in an input 1-block than with a fault which has previously acted on the second qubit in the same input 1-block; then, even though applying an i-decoder to the input 1-block yields the same state in both cases, applying i-decoders to the output of the 1-Rec in these two cases might yield different output states.

This observation tells us what direction we should take: we should look for an equivalence to faulty 0-Gas that *do* depend on the syndrome i-decoders find. To make the dependence on the syndrome information explicit, we will introduce a new version of i-decoders that we will call *coherent i-decoders*, or \mathcal{D} 's. Whereas i-decoders trace over all output ancillary qubits, coherent i-decoders do not; they maintain the *coherence* between the syndrome that ideal error-correction finds and the state of the decoded qubit. More specifically, let $|\psi\rangle_L$ denote some logical single-qubit state, and let $\{E_i\}$ denote the set of correctable Pauli errors of an $[[n, k, d]]$ stabilizer code; then the action of the coherent i-decoder is

$$\mathcal{D} : E_i |\psi\rangle_L \mapsto |\psi\rangle \otimes |e_i\rangle , \quad (3.5)$$

where $|\psi\rangle$ is the decoded single-qubit state carried, e.g., by the first output qubit, and $|e_i\rangle$ denotes the state of remaining $n-1$ qubits that records the syndrome.

If the states $\{E_i|\psi\rangle_L\}$ are a complete basis (i.e., if the code is perfect), then equation (3.5) completely characterizes the action of the coherent i-decoder. Otherwise, since the action in equation (3.5) is inner-product preserving, we may always choose some unitary extension of \mathcal{D} . Then, its inverse, \mathcal{D}^{-1} , is a *coherent i-encoder* that takes a qubit in the input state $|\psi\rangle$ and $n-1$ qubits in the state $|e_i\rangle$ and encodes them in the state $|\psi\rangle_L$ with Pauli error E_i . Of course, an i-decoder is just a coherent i-decoder that discards the $n-1$ output qubits carrying the syndrome information, or schematically:

$$\text{---} \boxed{\text{i-decoder}} \text{---} = \text{---} \boxed{\mathcal{D}} \text{---} \cdot \text{---} \times$$

This implies that lemma 1 in the previous chapter holds if i-decoders are replaced by coherent i-decoders: Indeed, lemma 1 is precisely the statement that the 1-Rec inside a good 1-exRec is correct *independent* of what syndrome the i-decoders find (since, otherwise, if there was a dependence on the syndrome, lemma 1 would not always hold independent of the syndrome that the i-decoders discard). For a 1-Rec simulating a single-qubit 0-Ga, lemma 1 now takes the form of the equality

$$\text{---} \boxed{1\text{-EC}} \text{---} \boxed{1\text{-Rec}} \text{---} \boxed{\mathcal{D}} \text{---} = \text{---} \boxed{1\text{-EC}} \text{---} \boxed{\mathcal{D}} \text{---} \boxed{\text{ideal } 0\text{-Ga}} \text{---} \cdot \text{---} \boxed{\mathcal{O}_{\text{syndr}}} \text{---}$$

where $\mathcal{O}_{\text{syndr}}$ is some operation acting on the syndrome.

Since coherent i-decoders are unitary, we are free to create pairs of coherent i-decoders and i-encoders (or simply, *decoder-encoder pairs*) via the identity $I = \mathcal{D} \cdot \mathcal{D}^{-1}$ which we can insert at any point during the quantum computation. In particular, we can insert decoder-encoder pairs preceding a bad 1-exRec in order to convert it to a faulty 0-Ga that depends on the syndrome that is processed by the coherent i-decoders. For a 1-exRec corresponding to a single-qubit 0-Ga, we can illustrate this as

$$\text{---} \boxed{\text{bad } 1\text{-exRec}} \text{---} \boxed{\mathcal{D}} \text{---} = \text{---} \boxed{\mathcal{D}} \text{---} \boxed{\text{faulty } 0\text{-Ga}} \text{---} \cdot \text{---}$$

where the faulty 0-Ga is defined as the operation

$$\text{---} \boxed{\text{faulty } 0\text{-Ga}} \text{---} = \text{---} \boxed{\mathcal{D}^{-1}} \text{---} \boxed{\text{bad } 1\text{-exRec}} \text{---} \boxed{\mathcal{D}} \text{---} \cdot \text{---}$$

It is important that decoder-encoder pairs are inserted on every 1-block preceding the *entire* bad 1-exRec because, in the next step, the coherent i-decoders will move to the left of the preceding

1-exRecs. We only want to consider these preceding 1-exRecs as bad if they are bad *independently* of the current 1-exRec. Thus, we insert our decoder-encoder pairs so that we *truncate* these preceding 1-exRecs from all 1-EC gadgets they share with the current bad 1-exRec. If the preceding truncated 1-exRecs are indeed bad independently of the current 1-exRec then, in the next step, we insert decoder-encoder pairs preceding them as well. And we repeat the same trick every time coherent i-decoders encounter a bad 1-exRec, thereby converting every independent bad 1-exRec to a faulty 0-Ga that is correlated with the syndrome.

On the other hand, every non-independent bad 1-exRec will, after truncation of the 1-EC gadgets it shares with its succeeding 1-exRecs, become good. Good truncated 1-exRecs are correct as can be easily seen as follows: First, assume i-decoders (instead of coherent i-decoders) are placed to the right of a truncated good 1-exRec. We can restore all truncated 1-ECs by pulling out ideal (i.e., faultless) 1-ECs from the i-decoders, since clearly

$$\boxed{\text{i-decoder}} = \boxed{\text{ideal 1-EC}} \boxed{\text{i-decoder}} .$$

Now, the good truncated 1-exRecs have become equivalent to full good 1-exRecs and lemma 1 applies. But, lemma 1 being true when phrased in terms of i-decoders implies that the correctness of good truncated 1-exRecs does not depend on the syndrome that the i-decoders find; hence, the lemma will also apply if coherent i-decoders are used instead. This shows that, in general, coherent i-decoders can be propagated to the left of the 1-Rec inside every good 1-exRec (truncated or not), thereby converting the 1-Rec to the ideal simulated 0-Ga applied on the decoded qubit(s) and some operation applied on the syndrome.

In effect, the syndrome acts as a common *bath* with which qubits at all faulty 0-Gas interact after level reduction: Consider a sequence of 1-exRecs that contains r independent bad 1-exRecs. By creating i-decoders, moving them to the left and finally annihilating them transforms the 1-exRecs to an equivalent circuit of 0-Gas with r faults; and these faults can be correlated since they correspond to the operation simulated by the bad 1-exRecs that depends on the syndrome. For instance, two bad 1-exRecs separated by a sequence of good intermediate 1-exRecs can be converted to two correlated faulty 0-Gas that are separated by ideal 0-Gas, or schematically:

$$\boxed{\text{bad 1-exRec}} \boxed{\text{good 1-exRecs}} \boxed{\text{bad 1-exRec}} \boxed{\mathcal{D}} = \boxed{\mathcal{D}} \boxed{\text{faulty 0-Ga}} \boxed{\text{ideal 0-Gas}} \boxed{\text{faulty 0-Ga}} .$$

It follows that even if noise is uncorrelated at the physical level, effective correlations between faults will arise as one considers the different coding levels in a recursive fault-tolerant simulation. Local noise has the property that it includes the possibility of such correlations and, for this reason,

local noise acting in the level- k simulation remains local in the equivalent level- $(k-1)$ simulation after level reduction and only the noise strength is transformed. In this sense, we may say that local noise is “stable” under level reduction.

3.3.2 The Combinatorics for Overlapping Rectangles

Finally, we need to justify why the bound (3.4) is valid even though some of the r specified bad 1-exRecs in the level- k simulation may overlap. We wish to express $F(\mathcal{I}_r^{(1)})$ as a sum of terms where for each term, each one of the r specified bad 1-exRecs contains from $t+1$ up to some maximum number of faults and it is bad *independently* from all other bad 1-exRecs. Because of the requirement for independence, the maximum number of faults in each of the r specified bad 1-exRecs can be either the number of locations in the full 1-exRec (if none of the 1-exRecs immediately following are bad), or the number of locations in the truncated 1-exRec (if one of the 1-exRecs immediately following is one of the other r specified bad 1-exRecs).

Nevertheless, for any specified fault path, whether each bad 1-exRec is truncated or not is unambiguously determined and, in order for a 1-exRec to be bad independently, it must contain at least $t+1$ faults. Therefore, we can still apply the bound (3.3) with the only change that for each fault path in the sum some 1-exRecs may be truncated, and which 1-exRecs are truncated varies from fault path to fault path so that the upper limits of each summation may vary. Since a truncated 1-exRec contains fewer locations than the full 1-exRec, the bound (3.3) applies in general independent of whether some of the r specified bad 1-exRecs overlap.

3.4 The Quantum Threshold Theorem for Local Noise

The virtue of the level-reduction lemma is that it is *composable*. We can apply it k times in succession to convert a level- k simulation to a level-0 simulation that produces the same output statistics; the level-0 simulation is directly implementing the desired quantum circuit but with a transformed noise strength. Consider first local stochastic noise of strength p acting on the initial level- k simulation. Then, by iterating k times the recursion equation $p^{(j)} \leq \binom{C}{t+1} (p^{(j-1)})^{t+1}$ with $p^{(0)} \equiv p$, the noise strength in the equivalent level-0 simulation is

$$p^{(k)} \leq p_{\text{thr}} \left(\frac{p}{p_{\text{thr}}} \right)^{(t+1)^k}, \text{ where } p_{\text{thr}} \equiv \left(\frac{C}{t+1} \right)^{-1/t}. \quad (3.6)$$

Hence, p_{thr} is the *threshold* noise strength, i.e., for $p < p_{\text{thr}}$ the *effective* noise that acts on the level- k simulation has strength that decreases doubly-exponentially with k as shown by equation (3.6). Using equation (1.15) with $p \rightarrow p^{(k)}$, we conclude that the error, δ , of the level- k simulation

is

$$\delta \leq 2L p_{\text{thr}} \left(\frac{p}{p_{\text{thr}}} \right)^{(t+1)^k}. \quad (3.7)$$

Therefore, any desired accuracy, $1 - \delta_0$, can be achieved by choosing k such that

$$(t+1)^k \geq \frac{\log(2L p_{\text{thr}}/\delta_0)}{\log(p_{\text{thr}}/p)}. \quad (3.8)$$

If c is the size (i.e., number of locations) of the largest 1-Rec and if d if the largest depth (i.e., number of time steps) in any 1-Rec then, because of the self-similarity of recursive simulations, all k -Recs will have size at most $c^k = ((t+1)^k)^{\log_{t+1} c}$ and depth at most $d^k = ((t+1)^k)^{\log_{t+1} d}$.

Similarly, for local noise that is not stochastic, if noise of strength ε acts on the initial level- k simulation, then the equivalent level-0 simulation is afflicted by local noise of strength

$$\varepsilon^{(k)} \leq \varepsilon_{\text{thr}} \left(\frac{\varepsilon}{\varepsilon_{\text{thr}}} \right)^{(t+1)^k}, \text{ where } \varepsilon_{\text{thr}} \equiv \left(\kappa \binom{C}{t+1} \right)^{-1/t}, \quad (3.9)$$

for $\kappa \geq e^{(C-t-1)\varepsilon_{\text{thr}}}$ provided $\varepsilon < \varepsilon_{\text{thr}}$; thus, ε_{thr} is the *threshold* noise strength. Similar to equation (2.15), the error, δ , of the level- k simulation can then be upper bounded by

$$\delta \leq 2eL\varepsilon^{(k)} \leq 2eL\varepsilon_{\text{thr}} \left(\frac{\varepsilon}{\varepsilon_{\text{thr}}} \right)^{(t+1)^k}. \quad (3.10)$$

Therefore, we arrive at a similar conclusion as for local stochastic noise, i.e., any desired accuracy, $1 - \delta_0$, can be achieved provided k is chosen so that

$$(t+1)^k \geq \frac{\log(2eL\varepsilon_{\text{thr}}/\delta_0)}{\log(\varepsilon_{\text{thr}}/\varepsilon)}. \quad (3.11)$$

With the help of the level-reduction lemma, we have thus proved the quantum threshold theorem for local noise that is stated next.

Theorem 4 (Quantum Accuracy Threshold for Local Noise). *Suppose gadgets can be constructed such that they satisfy properties 1 and 2 for a distance- d quantum error-correcting code, and let $t = \lfloor \frac{d-1}{2} \rfloor$. Let c be the largest size and d be the largest depth in any 1-Rec, and let C be the size of the largest 1-exRec. Suppose that noisy quantum circuits are afflicted by local noise of strength*

$$\varepsilon < \varepsilon_{\text{thr}} \equiv \left(\kappa \binom{C}{t+1} \right)^{-1/t},$$

where $\kappa \geq e^{(C-t-1)\varepsilon_{\text{thr}}}$ (or $p < p_{\text{thr}} \equiv \left(\frac{C}{t+1} \right)^{-1/t}$ if noise is stochastic). Then, for any fixed $\delta_0 > 0$, any ideal quantum circuit of size L and depth D can be simulated with error δ_0 by a noisy circuit of size $O(L(\log L)^{\log_{t+1} c})$ and depth $O(D(\log L)^{\log_{t+1} d})$.

As was mentioned in the statement of properties 2, an assumption in our construction of gadgets simulating measurements is that either classical processing is performed in encoded form using a classical error-correcting code or, if classical information is decoded to single classical bits, the classical gates that manipulate these bits are faultless. There is still a quantum accuracy threshold if the classical processing is noisy with noise strength p_{cl} , but the error, δ_0 , is limited to be of order p_{cl} in this case: one can perform fault-tolerant quantum error correction either by measuring the syndrome and doing the classical processing using a classical error-correcting code, or by performing the error recovery coherently using quantum operations and *without* quantum measurement. Then, the only obstacle is decoding the final outcome of the computation to the level of single classical bits, and the error of this decoding will be limited to be of order p_{cl} due to the noisy classical gates.

3.4.1 Improving the Threshold Estimate

In theorem 4, we estimated the quantum accuracy threshold by considering all possible ways, $\binom{C}{t+1}$, of placing $t+1$ faults in the largest 1-exRec. This estimate is however pessimistic since for some of these configurations of faults, the 1-Rec contained inside the 1-exRec will still be correct. Let us now discuss how this observation can help us obtain improved rigorous lower bounds on the quantum accuracy threshold.

Let us say that a set of locations inside a 1-exRec is *benign* if the 1-Rec contained in the 1-exRec is correct for arbitrary fault operators acting on all locations in the set. Naturally, if a set of locations is not benign, we will say it is *malignant*. Then, in this language, lemma 1 says that all sets of locations with cardinality at most t are benign; but we expect that many sets of locations with more than t faults will also be benign. We can generalize definition 7 accordingly.

Definition 8 (Goodness (revised and generalized)). *Consider a fault-tolerant quantum circuit simulation based on a distance- d quantum error-correcting code, and let $t = \lfloor \frac{d-1}{2} \rfloor$. A 1-exRec is bad if it contains faults at a malignant set of locations; if it is not bad it is good. Two bad 1-exRecs are independent if they are nonoverlapping or if they overlap and the earlier 1-exRec is still bad when the shared 1-ECs are removed.*

Then, lemma 1 is true by definition. Furthermore, with this generalized definition of goodness, lemma 3 will yield an improved estimate on the strength of noise that acts on the equivalent simulation that results after level reduction: Let us denote by A the number of malignant sets of $t+1$ locations and let us pessimistically consider that all sets of at least $t+2$ locations are malignant. Then, for local stochastic noise, the bound (3.1) can be replaced by the improved estimate

$$p^{(1)} \leq Ap^{t+1} + Bp^{t+2}, \quad (3.12)$$

where $B \equiv \binom{C}{t+2}$ is the number of sets of $t+2$ locations in the largest 1-exRec. Equivalently,

$$p^{(1)} \leq A' p^{t+1}, \quad (3.13)$$

which indicates an accuracy threshold, $p_{\text{thr}} \equiv (A')^{-1/t}$. Here, A' is obtained by solving the equation $A + B p_{\text{thr}} = A'$, or

$$A + B(A')^{-1/t} = A'. \quad (3.14)$$

For instance, for $t = 1$, we obtain $A' = \frac{1}{2}A \left(1 + \sqrt{1 + 4B/A^2}\right)$.

For local noise that is not stochastic, the sum of all fault paths with at least $t+2$ faults can be expressed as in lemma 2 with $s = t+2$. We need to add to this an additional contribution due to the sum of all fault paths with faults at *exactly* $t+1$ malignant coarse-grained locations; let us denote this term as $F(\mathcal{M}_{t+1})$. For each specified set of $t+1$ malignant coarse-grained locations, we place a fault operator at each of the locations in the set, and we place the ideal operations at all other locations inside the 1-exRec. Then, as in the proof of lemma 2, we expand each ideal operation as the unitary coupling with the bath *minus* a fault operator. In this derived expansion, each term with faults at m specific coarse-grained locations appears at most $\binom{m}{t+1}$ times (if every subset of $t+1$ out of the m coarse-grained locations is malignant) and may appear as few as zero times (if no subset of $t+1$ out of the m coarse-grained locations is malignant). On the other hand, each set of m faulty coarse-grained locations occurs at most $\binom{m-1}{t+1}$ times in the derived expansion arising from the original expansion of all fault paths with at least $t+2$ faults. Furthermore, these terms are opposite in sign to the terms that arise from the malignant sets of $t+1$ coarse-grained locations (the signs are $(-1)^{m-t-1}$ and $(-1)^{m-1-t-1}$ in the two cases). Therefore, after combining the two expansions, each fault path with m specific faults appears a number of times which is at most the larger of $\binom{m-1}{t+1}$ and $\binom{m}{t+1} - \binom{m-1}{t+1} = \binom{m-1}{t}$. Since, $\forall m \geq t+2$, $\binom{m-1}{t} \leq (t+1)\binom{m-1}{t+1}$, we can upper bound the sum over all fault paths with faults in all malignant sets of $t+1$ coarse-grained locations or faults in at least $t+2$ coarse-grained locations similar to equation (2.12):

$$\varepsilon^{(1)} \leq \|F(\mathcal{M}_{t+1}) + F(\mathcal{I}_{r \geq t+2})\|_{\text{sup}} \leq A\varepsilon^{t+1} + B\varepsilon^{t+2}, \quad (3.15)$$

where A is again the number of malignant sets of $t+1$ coarse-grained locations, and $B \equiv (t+1)\kappa\binom{C}{t+2}$ with $\kappa \geq e^{(C-t-2)\varepsilon}$. As before, we can write

$$\varepsilon^{(1)} \leq A' \varepsilon^{t+1}, \quad (3.16)$$

which implies $\varepsilon_{\text{thr}} \equiv (A')^{-1/t}$ (A' is again defined by equation (3.14)).

Overall, using the concept of malignant sets of locations, we have proved the following theorem.

Theorem 5 (Quantum Accuracy Threshold for Local Noise (revised)). *Suppose gadgets can be constructed such that they satisfy properties 1 and 2 for a distance- d quantum error-correcting code, and let $t = \lfloor \frac{d-1}{2} \rfloor$. Let c be the largest size and d be the largest depth in any 1-Rec, and let C be the size of the largest 1-exRec. Suppose that noisy quantum circuits are afflicted by local noise of strength*

$$\varepsilon < \varepsilon_{\text{thr}} \equiv (A')^{-1/t} ,$$

(or $p < p_{\text{thr}} \equiv (A')^{-1/t}$ if noise is stochastic) with A' as in equation (3.14) where A is the largest number of malignant sets of $t+1$ locations in any 1-exRec and $B \geq (t+1)e^{(C-t-2)\varepsilon_{\text{thr}}} \binom{C}{t+2}$ (or $B = \binom{C}{t+2}$ if noise is stochastic). Then, for any fixed $\delta_0 > 0$, any ideal quantum circuit of size L and depth D can be simulated with error δ_0 by a noisy circuit of size $O(L(\log L)^{\log_{t+1} c})$ and depth $O(D(\log L)^{\log_{t+1} d})$.

It is clear that we can obtain a further improvement in our estimate of the quantum accuracy threshold by resolving the sets of $t+2$ or more coarse-grained locations into malignant and benign; but we will not discuss such extensions here.

3.5 History and Acknowledgements

The analysis of recursive simulations and proofs of the quantum threshold theorem for independent stochastic noise were first given by Aharonov and Ben-Or [55], by Kitaev [30], and by Knill, Laflamme, and Zurek [52]; similar discussions were also given in [56] and [42]. All these proofs, with the exception of the discussion in [52] which is given without proof, applied to computation that is encoded using a code of distance at least 5. At the same time, most estimates for the quantum accuracy threshold were derived using the distance-3 Steane code [37]; this created an unsatisfactory situation since the formal proofs did not apply to this code.

Knill, Laflamme and Zurek in [52] sketched a method of analysis of fault-tolerant quantum computation based on a distance-3 code, but their argument only applied to level-1 simulations; in particular, it was not clear how to extend that argument to higher levels of a recursive simulation as is necessary for deriving an accuracy threshold condition. Starting in the spring of 2004, Gottesman, Preskill and I considered extending the argument in [52] to obtain a rigorous proof. Thinking in terms of the notions of goodness and correctness proved to be extremely beneficial in deriving the proof in [1] which is simple and, I think, elegant. The proof in [1] follows the tradition set by [55] and analyzes the accuracy of level- k gadgets using induction on k ; in [1] the inductive step was taken via a procedure we called *threshold dance* and which was suggested by Preskill. Shortly after [1], Reichardt also proved a quantum threshold theorem for distance-3 codes for the case of independent stochastic Pauli noise [57].

Encouraged by our success with distance-3 codes, Gottesman, Preskill and I subsequently started thinking about distance-2 codes motivated by the scheme of postselected quantum computation and the exciting numerical results of Knill [58, 59]. It was by thinking about the proof of the threshold theorem for postselected quantum computation that I understood how to obtain the proof of the quantum threshold theorem using a level-reduction procedure. In retrospect, Knill, Laflamme, and Zurek had the right intuition in [52] to consider level-1 simulations only: level reduction shows that recursive simulation can really be analysed one level at a time and so, in a certain sense, level 1 is all that matters!

Chapter 4

Further Applications of Level Reduction

4.1 Introduction

The proof of the quantum threshold theorem in the previous chapter was based on the level-reduction lemma. In essence, this lemma told us that recursive simulations can be analysed by considering how physical noise translates into an effective noise that acts on the encoded operations. Put differently, the concept of level reduction tells us that there are two questions to be addressed in order to prove the quantum threshold theorem: (i) Is our noise model general enough so that its structure is preserved under level reduction? And, if yes (as for local noise), (ii) how are the parameters characterizing this noise model (i.e., the noise strength for local noise) transformed under level reduction? The answer to the first question will tell us whether the quantum threshold theorem can be proved for a given noise model; the answer to the second question will tell us what the accuracy threshold condition is.

In this chapter, I will discuss how this intuition for the analysis of recursive simulations can be used to prove the quantum threshold theorem in two new settings. First, in §4.2, I will extend the proof to apply to local leakage noise; i.e., our noise model will include the possibility that a qubit may become inaccessible to our control operations by *leaking* to additional degrees of freedom beyond the two-dimensional Hilbert space ideally characterizing it. The proof will proceed by constructing gadgets that under a *leakage-reduction* procedure, transform local leakage noise to regular local noise for which the level-reduction lemma applies. Second, in §4.3, I will extend the proof to measurement-based models of quantum computation where elementary quantum gates are themselves simulated by gadgets. In this case again, the proof will proceed by showing that under a suitable *model-conversion* procedure, noisy measurement-based quantum computation is equivalent to noisy quantum circuit computation for which the level-reduction lemma applies.

4.2 Leakage Noise

In many physical implementations, a qubit is a two-dimensional *subspace* of a higher-dimensional Hilbert space. Most usually this happens when a qubit is simply two low-lying levels in a many-level physical system; examples include an ion-trap qubit, an optical qubit which is the dual-rail subspace of two photonic modes [60], a Josephson-junction qubit which is formed by the lowest energy states in a double well potential [61], and a quantum-dot qubit which is encoded in electron-spin states [62, 63]. For such qubits, *leakage faults* transfer amplitude in and out of the two-dimensional subspace defining the qubit and are likely to be an important source of errors. In this section, I will refer to the two-dimensional qubit space as the *system space*, and to its complement—i.e., the space occupied by qubits that have leaked out of the system space—as the *leakage space*.

The first protection against leakage faults is the use of low-decoherence qubit encodings and dynamical decoupling methods (as analyzed, e.g., in [64] and [65]). After these methods are applied to reduce the occurrence of leakage, quantum error correction will need to be used to protect against the remaining rare leakage faults. However, the problem is that standard quantum error-correcting codes are not designed to correct errors due to leakage faults directly since the correctable error operators are typically only Pauli operators with support on the system space. Thus, before quantum error correction is applied, we first need to *convert* errors due to leakage faults to regular errors that occur in the system space.

Converting qubits that have leaked out of the system space to erroneous qubits *in* the system space can be achieved in two ways: One way is to detect leakage and replace any leaked qubits by *new* qubits in the system space. For instance, in optical schemes for quantum computation, parity measurements of the photon occupation number in different modes can be performed that will indicate whether photon loss has occurred [60]. Another example is the indirect detection of leakage for ion-trap qubits by observing the emission of photons from a cavity [66]. Some leakage detection is also possible in solid-state schemes where qubits are encoded in two electron-spin states of a double quantum dot [63]. More generally, leakage can be detected by running each qubit through an appropriate leakage-detection circuit whose construction depends on the nature of the leakage space and the implementation of quantum gates. For instance, in implementations where the CNOT gate acts trivially if its control qubit is *not* in the system space, leakage detection can be achieved by the circuit in figure 4.1 [42, 67].

Let us refer to any procedure for detecting leakage as a *leakage-detection unit*, or *LDU*. If leakage detection specifically points to one qubit, we can simply replace the qubit by a new qubit initialized in some fixed state. But, in general, LDUs may act on multiple qubits and may not give us specific information about which qubit has leaked (e.g., in ion-trap schemes, detecting a photon escaping from the cavity indicates that one of the qubits inside the trap leaked but not which one). If leakage

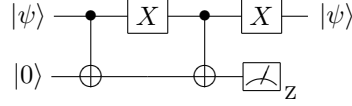


Figure 4.1. A circuit for detecting leakage in implementations where the CNOT gate acts trivially if its control qubit is in the leakage space. To test a qubit for leakage, we use an ancillary qubit initialized in the state $|0\rangle$ and a sequence of two CNOT and two bit-flip, X , gates. If the qubit is in the system space, then its state is preserved but the state of the ancillary qubit is flipped to the state $|1\rangle$; hence, the measurement gives outcome -1 . Otherwise, the ancillary qubit remains in the state $|0\rangle$ if leakage has occurred and the measurement gives outcome $+1$.

is detected on ancillary qubits, we can throw away all qubits and repeat the ancilla preparation circuit anew some fixed number of times. Otherwise, if leakage is detected on data qubits, we can use quantum teleportation to transfer the state of each qubit to a new qubit in the system space.

Alternatively to using leakage detection, we can constantly refresh qubits so that any leaked ones will be quickly replaced by erroneous qubits in the system space without us detecting whether leakage occurred; let us refer to such a procedure as a *leakage-reduction unit*, or *LRU*. The quantum teleportation circuit in figure 1.2 is a natural implementation of an LRU [68]. Indeed, if no leakage fault occurs during the execution of the teleportation circuit, the state of the third output qubit will always be in the system space independent of whether the first input qubit had leaked out of its system space or not. Hence, quantum teleportation offers a general solution to leakage reduction for cases when leakage detection, which depends on the specific implementation and the nature of the leakage space, is either limited or difficult to implement.

The remaining of this section is organized as follows: First, in §4.2.1, I describe a Hamiltonian local leakage noise model that generalizes the Hamiltonian noise model discussed in §2.2. In §4.2.2, I describe how quantum circuits need to be modified to protect against leakage faults and I prove the central result of this analysis, the *leakage-reduction lemma*. In §4.2.3, I use the leakage-reduction lemma together with the level-reduction lemma from the previous chapter to prove the quantum threshold theorem for local leakage noise. Finally, in §4.2.4, I discuss conditions for where LRUs must be placed inside a quantum circuit and where they can be omitted.

4.2.1 The Local Leakage Noise Model

Let \mathcal{H}_S be the system Hilbert space of all qubits, with $\mathcal{H}_{S[i]}$ the Hilbert space of qubit i . There is an extension of each $\mathcal{H}_{S[i]}$, $\mathcal{H}_{S_{ext}[i]} = \mathcal{H}_{S[i]} \oplus \mathcal{H}_{L[i]}$, so that $\mathcal{H}_{L[i]}$ is the leakage space of qubit i . Furthermore, we will assume that $\mathcal{H}_L = \otimes_i \mathcal{H}_{L[i]}$, i.e., the leakage spaces of different qubits are disjoint; this condition is fulfilled in most physical systems. Similar to §2.2, we will assume the following Hamiltonian

$$H = H_S + H_B + H_{SB} + H_{\text{leak}} , \quad (4.1)$$

where the additional leakage part of the Hamiltonian is of the form

$$H_{\text{leak}}(t) = H_{SL(B)}(t) + H_L(t) . \quad (4.2)$$

Here, H_L describes the arbitrary evolution in the leakage spaces and any coupling of the leakage spaces to the bath, and the time-dependent $H_{SL(B)}$ is a linear combination of operators coupling the leakage and system spaces and, potentially, the bath, B . An example of leakage that includes a coupling to a bath is a trapped ion in a cavity where transitions into or out of the two-level subspace can create or annihilate photons in the cavity. We will take $H_{SL(B)}$ to be of the form

$$H_{SL(B)}(t) = \sum_j H_{SL(B),j} , \quad (4.3)$$

where $H_{SL(B),j}$ only couples the system and leakage spaces of the set of qubits on the support of location j in the ideal quantum computation.

This noise model can be analysed using time-resolved fault paths as in §2.2. Let us assume that for all noisy locations the system-bath coupling for regular faults has the form of equation (2.3) and satisfies condition (2.6), and let us also make a similar assumption for the coupling between the system and leakage spaces, $\|H_{SL(B),j}\|_{\text{sup}} \leq \lambda_1$. Then, we can conclude that the sum of all fault paths with faults at r specific coarse-grained locations, \mathcal{I}_r , has sup norm $\|F(\mathcal{I}_r)\|_{\text{sup}} \leq ((\lambda_0 + \lambda_1)\tau)^r$. Hence, our leakage noise model is local with strength $\varepsilon \equiv (\lambda_0 + \lambda_1)\tau$.

4.2.2 The Leakage-Reduction Lemma

Before proceeding further, at this point it will be helpful to pause and discuss in more detail why leakage faults require a separate fault tolerance analysis. What distinguishes leakage faults from regular qubit faults is that once a qubit has leaked, its subsequent self-evolution or interaction with other qubits will depend on the details of the leakage space which vary in different implementations. In particular, if one of the input qubits to a multi-qubit quantum gate has leaked, the gate will in principle operate on the extended space of *all* its input qubits even if it is executed ideally. Therefore, leakage errors cannot be propagated in a simple way through ideal gates since this propagation depends on the particular gate implementation, i.e., the specification of how gates operate on the extended space. Since the design of fault-tolerant quantum circuits is guided by the particular ways in which errors propagate through ideal gates, a circuit designed to be robust against regular qubit faults will not, in general, be effective in maintaining fault tolerance against leakage.

To ensure that all faultless 0-Gas in a quantum circuit act in the system space of their inputs, we will consider inserting an LRU before every *gate* 0-Ga. We need not place LRUs preceding measurement 0-Gas which have only classical output or preceding qubit-preparation 0-Gas which

have no quantum input. When an LRU is executed without a fault, it guarantees that (i) if its input is in the system space then the identity operation is applied, and (ii) if its input is in the leakage space then the output qubit state is in the system space. As discussed above, the quantum teleportation circuit in figure 1.2 is a natural way to satisfy these two requirements. If leakage detection per qubit is possible, we can alternatively place an LDU before every gate 0-Ga. An LDU outputs a classical bit indicating whether leakage has occurred, in which case the leaked qubit can be replaced by a new qubit in some fixed state in the system space. In the discussion that follows, we will assume LRUs precede gate 0-Gas, but the idea is essentially the same if LDUs are used instead.

Given a quantum circuit with LRUs before every gate 0-Ga, it is clear that *if* the LRUs are faultless, this circuit subjected to leakage faults is equivalent to the circuit without LRUs subjected to regular qubit faults. This is because LRUs replace any leaked input with a state in the system space before the next 0-Ga is applied. However an LRU can fail. It can fail due to a regular qubit fault, but it can also fail due to a leakage fault; e.g., if LRUs are implemented by quantum teleportation, there can be a leakage fault acting on the outgoing qubit. The essential insight for analyzing leakage is that an LRU implements leakage correction at the physical level, *level 0*, just as a k -EC gadget performs regular error correction at level k of a recursive simulation. The leakage-reduction lemma that follows formalizes this intuition: This lemma allows us to transform a quantum circuit where level-0 leakage correction is performed by LRUs to a quantum circuit where LRUs are removed and the effective noise only includes regular qubit faults.

But we first need some additional definitions; let us start by defining *level-0 rectangles*, or *0-Recs*. For quantum circuits without LRUs this would be trivial: a 0-Rec is just a 0-Ga and it is bad (and incorrect) when it is faulty. When LRUs are inserted in a quantum circuit, for every gate 0-Ga we define a 0-Rec as the 0-Ga preceded by LRUs on all its inputs (Fig. 4.2); for a qubit-preparation or measurement 0-Ga, the 0-Rec coincides with the 0-Ga. We will also say that a 0-Rec *simulates* the 0-Ga that remains after the LRUs are removed.

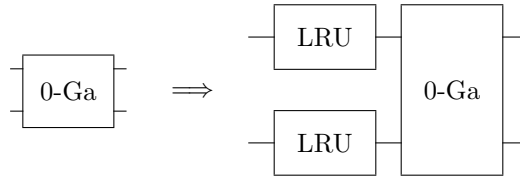


Figure 4.2. To combat leakage faults, we insert LRUs on the inputs of every gate 0-Ga. The combination of a 0-Ga with its preceding LRUs we call a *0-Rec*.

Similar to definition 5, we would like to define what it means for 0-Recs to be correct. In this new definition, the role of an ideal decoder will now be played by an *ideal LRU*, or *i-LRU*. This device can be thought of as an ideal (i.e., faultless) execution of quantum teleportation where the classical

bits are discarded after any Pauli correction operators are applied; just as i-decoders, i-LRUs are tools for our analysis and do not represent physical operations applied during the computation. The definition of correctness and goodness for 0-Recs is as follows.

Definition 9 (Correctness for 0-Recs). *A 0-Rec is correct if the 0-Rec followed by i-LRUs is equivalent to i-LRUs followed by the ideal 0-Ga that the 0-Rec simulates.*

Definition 10 (Goodness for 0-Recs). *A 0-Rec is good if it contains no leakage or regular faults, otherwise it is bad.*

Since a good 0-Rec contains no faults, its LRUs operate ideally; hence, the following lemma holds.

Lemma 4. *A good 0-Rec is correct.*

Definition 10 also applies to preparation and measurement 0-Recs which do not contain any LRUs: these 0-Recs are also good when they are faultless. Moreover, the correctness of a good measurement 0-Rec implies that it can be replaced by an i-LRU followed by the corresponding ideal measurement 0-Ga providing a means to create i-LRUs (see §4.2.2.1). Similarly, correctness implies that a good preparation 0-Rec followed by an i-LRU can be replaced by the corresponding ideal preparation 0-Ga alone, thus annihilating the i-LRU.

We can now state our basic lemma.

Lemma 5 (Leakage Reduction). *Consider any quantum circuit where LRUs have been inserted preceding every gate 0-Ga, and let C_{leak} be the maximum number of locations in any 0-Rec. Then, if this circuit is subject to local leakage noise of strength ε , there is a quantum circuit with the LRUs removed that produces the same output probability distribution and is subject to regular local noise with strength*

$$\tilde{\varepsilon} \leq C_{\text{leak}} \varepsilon .$$

Proof. The proof is similar to the proof of the level-reduction lemma. We imagine first creating i-LRUs out of measurement 0-Recs thereby transforming them to measurement 0-Gas. Next, we imagine propagating these i-LRUs to the left through gate 0-Recs thereby transforming them to gate 0-Gas. Finally, we imagine annihilating the i-LRUs inside preparation 0-Recs thereby transforming them to preparation 0-Gas. By lemma 4, every good 0-Rec will be thus replaced by the ideal simulated 0-Ga; in §4.2.2.1, we will also show that every bad 0-Rec can be transformed to a faulty 0-Ga that operates on the system space. Hence, with this maneuver which we can visualize as an “i-LRU wave” propagating from the right to the left of the quantum circuit which includes LRUs, we will obtain a circuit with the LRUs removed that realizes exactly the same computation, and this circuit is subject to regular qubit faults alone.

Consider some set, $\mathcal{I}_r^{(0)}$, of r specific locations in this equivalent circuit. In order for these locations to be faulty, the corresponding 0-Recs in the initial circuit must all be bad, i.e., they must each contain at least one fault. Then, by using a similar counting argument as in the proof of the level-reduction lemma, we can upper bound the sup norm of the sum of all fault paths with faults at all 0-Recs corresponding to the locations in $\mathcal{I}_r^{(0)}$ by

$$\|F(\mathcal{I}_r^{(0)})\|_{\text{sup}} \leq (\kappa_0 C_{\text{leak}} \varepsilon)^r, \quad (4.4)$$

where $\kappa_0 \geq e^{(C_{\text{leak}}-1)\varepsilon}$. In fact, it is sufficient to take $\kappa_0 = 1$ as can be seen by repeating the analysis of time-resolved fault paths in §2.2 and considering each 0-Rec as *one* operation that takes time $C_{\text{leak}} \tau$. We conclude that the noise acting on the equivalent circuit is local with strength $\tilde{\varepsilon} \leq C_{\text{leak}} \varepsilon$.

This proves the lemma. □

There is a corresponding threshold condition for the case when local leakage noise is stochastic. However, it is important to emphasize that in most physical settings, leakage is an inherently non-Markovian process, in particular when it does not involve interactions with a bath which can induce decoherence between different fault paths. In cases when leakage faults can be assigned a probability, p , the same analysis holds where now the noise strength afflicting the equivalent circuit will be $\tilde{p} \leq C_{\text{leak}} p$.

4.2.2.1 Bad Level-0 Rectangles As Simulated Faults

It remains to explain how i-LRUs can be moved to the left of bad 0-Recs thereby transforming them to faulty 0-Gas that act on the system space. The idea for accomplishing this is essentially the same as for converting bad 1-exRecs to simulated faulty 0-Gas as discussed in §3.3.1.

The first step is to consider the coherent version of an i-LRU. Like i-decoders, i-LRUs generate a syndrome; taking i-LRUs to be ideal quantum teleportation circuits, the syndrome consists of the measurement bits. We can then define a coherent invertible teleportation in which one, instead of measuring, performs CNOT and CPHASE gates on the output qubit (Fig. 4.3). Furthermore, it is possible to *define* the CNOT and CPHASE gates in this i-LRU circuit to act trivially on their target qubits when their control qubits are in the leakage space. Then, the output qubit will always be in the system space independent of whether the input qubit has leaked or not; leakage will be confined to the *syndrome qubits*, i.e., the first two output qubits in figure 4.3. Let us denote a coherent i-LRU as \mathcal{LRU} , so that its inverse is denoted as \mathcal{LRU}^{-1} . Hence, for our analysis, we can create *LRU pairs*, $\mathcal{LRU}^{-1} \cdot \mathcal{LRU} = I$, and insert them at any point during the quantum computation.

Consider first those qubits in a quantum circuit that are measured. A measurement can be

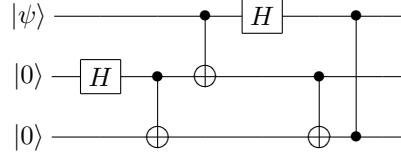


Figure 4.3. A coherent version of quantum teleportation that can, e.g., be used to construct our coherent i-LRUs. Instead of measuring the first two qubits, we apply the Pauli corrections coherently via a Hadamard rotation followed by a CNOT and a CPHASE gate. If we define the CNOT and CPHASE gates in this circuit to act trivially if their control qubits are in the leakage space, then independent of the state of the input qubit, the third qubit is always in some state in the system space.

viewed as an \mathcal{LRU} followed by another measurement whose input is always in the system space: We can insert an LRU pair before every measurement 0-Ga, \mathcal{M} , and identify $\mathcal{M}' = \mathcal{M} \cdot \mathcal{LRU}^{-1}$ as a measurement 0-Ga that always acts on the system space (and depends on the syndrome qubits input to the \mathcal{LRU}^{-1}). We can next propagate the \mathcal{LRUs} of each LRU pair to the left through good 0-Recs transforming them to the ideal simulated 0-Gas. Indeed, since lemma 4 applies to i-LRUs that measure the syndrome qubits and finally discard the syndrome information, the same is also true for coherent i-LRUs and do not measure the syndrome qubits. When we encounter a bad 0-Rec, we insert an LRU pair preceding it in all its inputs. The leading \mathcal{LRU} in each LRU pair can then be moved further to the left through the remaining 0-Recs until we encounter another bad 0-Rec and repeat by inserting new LRU pairs preceding it.

Furthermore, a bad 0-Rec grouped together with the \mathcal{LRU}^{-1} or \mathcal{LRU}^{-1} s preceding it and the \mathcal{LRU} or \mathcal{LRUs} succeeding it can be interpreted as a faulty 0-Ga that acts in the system space. For example, consider a 0-Rec simulating a single-qubit 0-Ga. The input to the \mathcal{LRU}^{-1} is a qubit always in the system space and some input syndrome qubits. Similarly, the \mathcal{LRU} outputs a qubit always in the system space and some syndrome qubits. Hence, the combined operation $\mathcal{LRU} \cdot \text{0-Rec} \cdot \mathcal{LRU}^{-1}$ can be viewed as some 0-Ga supported in the system space alone. For a good 0-Rec, this 0-Ga is exactly the 0-Ga simulated by the 0-Rec—this is our correctness property. What we have shown is that even bad 0-Recs can be viewed as simulating some faulty 0-Ga that is supported on the system space; and different faulty 0-Gas may be *correlated* as they depend on the syndrome processed by the LRU pairs.

4.2.3 The Quantum Threshold Theorem for Local Leakage Noise

The leakage-reduction lemma states that local leakage noise is effectively transformed to regular local noise if we insert LRUs preceding every gate 0-Ga in a quantum circuit. We will now consider inserting LRUs preceding every gate 0-Ga in a quantum circuit, \mathcal{C}_{rec} , that realizes a recursive fault-tolerant simulation designed to be robust against regular local noise as described in chapter 3; let

the resulting quantum circuit be denoted as $\mathcal{C}_{\text{rec}}^{\text{LRU}}$. Then, the leakage-reduction lemma tells us that local leakage noise of strength ε acting on $\mathcal{C}_{\text{rec}}^{\text{LRU}}$ is equivalent to regular local noise of strength at most $C_{\text{leak}}\varepsilon$ acting on \mathcal{C}_{rec} . Applying theorem 5 to \mathcal{C}_{rec} completes the proof of the quantum threshold theorem for local leakage noise which is stated next.

Theorem 6 (Quantum Accuracy Threshold for Local Leakage Noise). *Suppose gadgets can be constructed such that they satisfy properties 1 and 2 for a distance- d quantum error-correcting code, and let $t = \lfloor \frac{d-1}{2} \rfloor$. Let c be the largest size and d be the largest depth in any 1-Rec. Let C be the size of the largest 1-exRec, and C_{leak} be the size of the largest 0-Rec. Suppose that noisy quantum circuits are afflicted by local leakage noise of strength*

$$\varepsilon < \varepsilon_{\text{thr}} \equiv (A')^{-1/t} / C_{\text{leak}} ,$$

(or $p < p_{\text{thr}} \equiv (A')^{-1/t} / C_{\text{leak}}$ if noise is stochastic) with A' as in equation (3.14) where A is the largest number of malignant sets of $t+1$ faults in any 1-exRec and $B \geq (t+1)e^{(C-t-2)\varepsilon_{\text{thr}}} \binom{C}{t+2}$ (or $B = \binom{C}{t+2}$ if noise is stochastic). Then, for any fixed $\delta_0 > 0$, any ideal quantum circuit of size L and depth D can be simulated with error δ_0 by a noisy circuit of size $O(L(\log L)^{\log_{t+1} c})$ and depth $O(D(\log L)^{\log_{t+1} d})$.

The value of C_{leak} depends on our method for realizing LRUs. If r is the number of elementary operations in an LRU, then $C_{\text{leak}} \leq 2r + 1$ since we can restrict to 0-Recs simulating two-qubit 0-Gas and such 0-Recs contain two LRUs. In particular, if LRUs are implemented by the quantum teleportation circuit in figure 1.2, then $r = 7$ and the accuracy threshold for local leakage noise is diminished by at most an order of magnitude compared to the accuracy threshold for regular local noise. However, this conclusion is overly pessimistic since it is based on our assumption that LRUs are inserted preceding every gate 0-Ga. Although this assumption is helpful for proving theorem 6, it is not necessary. In the next section, we will discuss conditions for when LRUs can be omitted while still maintaining protection against leakage.

4.2.4 Contracting Level-0 Rectangles

Our motivation for omitting LRUs between successive 0-Gas is twofold: First, eliminating unnecessary LRUs will make the fault-tolerant circuits more efficient. And, second, we expect that this will help increase the accuracy threshold for leakage noise bringing it closer to the threshold for regular noise. The idea of omitting LRUs also inspires a similar modification in fault-tolerant circuits designed to protect against regular qubit errors. In that case, it may be similarly advantageous to omit error-correction gadgets; we will return to this point in the next chapter.

Consider the union of two or more consecutive 0-Recs with any intermediate LRUs omitted. We will call such an object a *contracted level-0 rectangle* or *0-conRec*; an example is shown in figure 4.4.

If this 0-conRec contains no faults, then the initial LRUs will operate ideally and the 0-conRec will be correct. In this case, the omission of the intermediate LRU has no consequence. However, assume instead that a leakage fault occurred in the first 0-Ga inside the 0-conRec only. Then we cannot interpret the second 0-Ga as an ideal 0-Ga even though it is executed without fault. This is because if the input to a 0-Ga is not in the system space, then the action of the 0-Ga can in principle be arbitrary in the extended space of *all* its inputs. This shows that in general, if the first 0-Ga inside a 0-conRec fails, all subsequent 0-Gas inside the 0-conRec will fail as well.

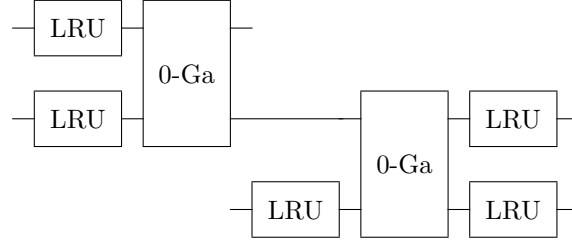


Figure 4.4. An example of a 0-conRec obtained by contracting two successive 0-Recs that simulate two two-qubit 0-Gas.

In general, lemma 4 applies also to 0-conRecs; i.e., a 0-conRec without faults followed by i-LRUs is equivalent to i-LRUs followed by the ideal sequence of 0-Gas the 0-conRec simulates. On the other hand, if a 0-conRec is bad, it can be viewed as implementing the same sequence of 0-Gas but with all 0-Gas potentially faulty¹.

Consider finally a recursive simulation where LRUs have been inserted preceding every gate 0-Ga to protect against leakage. Let us imagine we have contracted r successive 0-Recs to form a 0-conRec, and also let $\mathcal{I}_r^{(0)}$ be the set of locations that are simulated by this 0-conRec. Then, this 0-Rec contraction is allowed if no subset of $\mathcal{I}_r^{(0)}$ forms a malignant set inside any 1-exRec. For instance, a case where contracting 0-Recs is allowed is in a sequence of 0-Recs simulating single-qubit gate or memory 0-Gas: in this case, it suffices to keep the LRU at the beginning and omit all others, since any sequence of single-qubit faulty 0-Gas is no worse than just one fault in any of these 0-Gas. Another example is a 0-Rec simulating a preparation 0-Ga followed by a 0-Rec simulating a two-qubit gate 0-Ga, where again the intermediate LRU can be omitted².

4.3 Measurement-Based Quantum Computation

In this section, I will discuss fault-tolerant quantum computation in cases where every gate in an ideal quantum circuit is *simulated* by a sequence of quantum measurements on an larger Hilbert space that includes ancillary qubits. I will refer to such schemes where the basic primitive is quantum

¹This can be achieved by inserting LRU pairs inside the bad 0-conRec as in §4.2.2.1.

²Then, note that the resulting 0-conRec will contain no LRUs.

measurement as models of *measurement-based* quantum computation. Although several such models were initially proposed [69, 70, 71, 72], it has become clear that they are all related to one another [73, 74, 75, 76] and, moreover, they can all be understood in terms of a simple computation primitive which is known as *gate teleportation*.

The concept of gate teleportation was developed to address the problem of constructing a quantum universal set of gate gadgets satisfying properties 2 [51, 42, 32, 77]. On the other hand, the interest in measurement-based quantum computation stems primarily from questions relating to physical implementations. In particular, computation in measurement-based models proceeds by preparing suitable ancillary states *off-line*, i.e., separate to the main computation and, after preparation is complete, performing a joint quantum measurement on the ancillary qubits and some data qubits that is equivalent to applying a quantum gate on the latter. Thus, since some of the complexity of performing quantum gates is moved to the off-line preparation of ancillary states, measurement-based models are beneficial in those experimental situations where the direct implementation of quantum gates is problematic (see, e.g., [78]).

The rest of this section is structured as follows: In §4.3.1, I discuss the concept of gate teleportation and how it relates to measurement-based quantum computation. Then, in §4.3.2, I will prove the *model-conversion* lemma which shows that noisy measurement-based quantum computation can be analyzed in terms of an equivalent noisy quantum-circuit computation. This will then naturally lead to a proof of the threshold theorem for local noise for measurement-based quantum computation which I give in §4.3.3.

4.3.1 Gate Teleportation and Quantum Computation by Measurements

Quantum teleportation can be thought of as the answer to the question: Can we transfer the state of one qubit to another qubit?³ The answer is affirmative [27] as illustrated by the circuit in figure 1.2: It suffices to prepare two ancillary qubits in a *Bell state*, $|\Phi_0\rangle \propto |00\rangle + |11\rangle$, perform a CNOT gate from our first qubit to one of the two ancillary qubits and measure both our first qubit and this ancilla. The second ancillary qubit then acquires the state of the first qubit, $|\psi\rangle$, up to Pauli operators that depend on the outcomes of the two measurements. The entangled state $|\Phi_0\rangle$ is necessary if the qubits carrying the unknown state $|\psi\rangle$ in the beginning and in the end cannot interact directly. Else, simpler quantum teleportation circuits exist [77] as, e.g., the one in figure 4.5.

This circuit can alternatively be viewed as implementing a *teleportation* of the Hadamard gate. That is, instead of looking at this circuit as transferring the state $|\psi\rangle$ from the first to the second qubit, we can instead consider the Hadamard gate as being teleported. In general, the term *gate teleportation* denotes any procedure that achieves the simulation of some quantum gate using ancil-

³We demand that we do not know the state to be transferred because, if we did, then the protocol is trivial: we could just prepare a second qubit in that exact state.

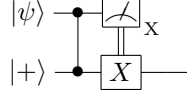


Figure 4.5. The output of the circuit is $H|\psi\rangle$; i.e., the input state has been *teleported* to the second qubit up to the application of a Hadamard gate. Recall that $|+\rangle \propto |0\rangle + |1\rangle$ is the +1 eigenstate of the Pauli operator X .

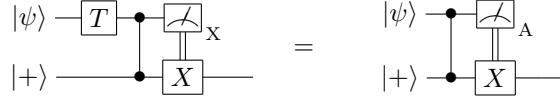


Figure 4.6. On the left, the T gate is first applied and, then, the circuit in figure 4.5 is executed; hence, the output is $HT|\psi\rangle$. On the right, an equivalent circuit where the T gate has been removed and the measurement is now performed along the eigenbasis of the operator $A \equiv T^\dagger X T$ [71, 72]; again, the output is $HT|\psi\rangle$, i.e., the circuit *teleports* the gate HT .

lary states and gates which are, depending on the context, easier to implement [51, 42, 32, 77]. In chapter 5, we will discuss how the idea of gate teleportation is used to obtain a universal set of gate gadgets satisfying properties 2; in this context, the gates that are easy to implement belong to the Clifford group. In measurement-based models [70, 69, 72], the set of easy to implement operations includes some fixed entangling gate (e.g., the CNOT or CPHASE gate), the initialization of qubits in fixed Pauli eigenstates (e.g, the +1 eigenstates of Z or X , $|0\rangle$ and $|+\rangle$ respectively) and single-qubit or two-qubit measurements along arbitrary orthonormal bases.

More specifically, consider the *graph-state model* of quantum computation [79] where the set, \mathcal{O} , of available operations is $\mathcal{O} = \{\text{CPHASE}, \mathcal{P}_{|+\rangle}, \mathcal{M}_A\}$; here, $\mathcal{P}_{|+\rangle}$ denotes the preparation of a qubit in the state $|+\rangle$, and \mathcal{M}_A denotes the measurement of a qubit along the eigenbasis of *any* single-qubit observable A . To show that \mathcal{O} is universal for quantum computation, it suffices to show that any gate from the universal set $\mathcal{G} = \{\text{CNOT}, H, T\}$ can be simulated (see proposition 1 in chapter 1). Indeed, the circuit in figure 4.5 shows how to simulate the Hadamard gate and, then, the CNOT can be simulated as $\text{CNOT} = (I \otimes H) \text{CPHASE} (I \otimes H)$. It remains to simulate the $T = \exp(-i\frac{\pi}{8}Z)$ gate. The circuit in figure 4.6 shows how to implement the simulation of HT using a measurement of the observable $T^\dagger X T$; then, T can be simulated as $H \cdot HT = T$.

Since the only unitary interactions available in measurement-based models are entangling gates such as the CPHASE gate, the Pauli operators that are conditioned on the measurement outcomes within each gate simulation cannot be directly applied; we then say that each gate is simulated up to *known* Pauli *corrections*. This is not a problem for the quantum computation as a whole, since we can easily propagate Pauli operators through subsequent gates—e.g., the CPHASE gate belongs in the Clifford group and, so, conjugates Pauli operators to Pauli operators—and we can eventually absorb the Pauli corrections inside subsequent measurements similar to the trick in figure 4.6. In

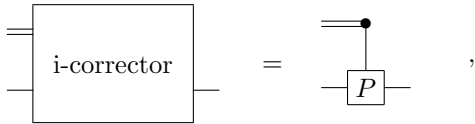
this way, our knowledge of the Pauli corrections allows us to adapt some of the bases of subsequent measurements in order for the desired overall simulation to be executed.

4.3.2 The Model-Conversion Lemma

At first, it seems that the analysis of noisy measurement-based quantum computation is complicated by the fact that the measurement bases within each gate simulation depend on the outcomes of previous measurements. For example, if one measurement outcome is erroneous, then this will lead to adapting incorrectly some future measurement bases; hence, one fault can cause a series of other *correlated* faults. In order to obtain a proof of the quantum threshold theorem for measurement-based models, we will therefore need to show that such correlated faults are not an obstacle to executing fault-tolerant quantum computation.

Before proceeding further, it is helpful to develop a formalism of describing individual gate simulations in measurement-based models and how simulations compose together. For every simulated gate, let us define a *level-0 simulator*, or *0-Sim*, as the union of operations required for simulating this gate. For example, the 0-Sim corresponding to the Hadamard gate consists of the preparation of an ancillary qubit in the state $|+\rangle$, a CPHASE gate and a measurement along the eigenbasis of X (Fig. 4.5)—note that the conditional gate X is not contained in the 0-Sim since Pauli corrections are never directly applied.

As discussed previously, every gate simulation in measurement-based models is achieved up to known Pauli corrections that are recorded in a classical memory and are used to adapt future measurement bases. To make the role of Pauli corrections more transparent, we will define an *ideal Pauli corrector*, or *i-corrector*. An i-corrector is an ideal (i.e., faultless) device that conditioned on the classical record of the Pauli correction of a given qubit, applies this Pauli correction operator and discards the classical record. Like i-decoders and i-LRUs, i-correctors will be tools for our analysis and do not correspond to physical operations applied during a measurement-based computation. If we use a double line to denote the *classical* bits carrying the information about the Pauli correction of some qubit, then the i-corrector acting on this qubit is the operation

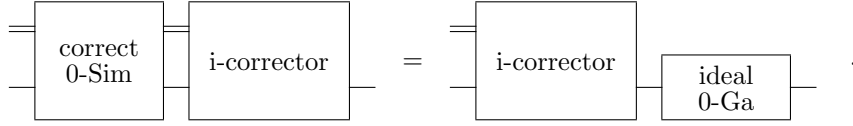


where P denotes the Pauli correction operator that the classical record indicates.

We can now state what it means for a 0-Sim to correctly simulate some 0-Ga.

Definition 11 (Correctness for 0-Sims). *A 0-Sim is correct if the 0-Sim followed by i-correctors is equivalent to i-correctors followed by the ideal 0-Ga that the 0-Sim simulates. For a single-qubit gate*

0-Ga, correctness means schematically:



We will also say that

Definition 12 (Goodness for 0-Sims). *A 0-Sim is good if it contains no faults, otherwise it is bad.*

Since a good 0-Sim contains no faults, it operates ideally; hence, the following lemma holds.

Lemma 6 (0-Sim-Correct). *A good 0-Sim is correct.*

It is sufficient to construct 0-Sims for every gate 0-Ga in a finite universal gate set; e.g., such constructions for the graph-state model were discussed in the previous section. Definitions 11 and 12 also hold for 0-Sims simulating qubit preparation or measurement 0-Gas. The correctness of good measurement 0-Sims allows us to replace them by i-correctors followed by the corresponding ideal measurement 0-Gas, thereby creating i-correctors. Similarly, the correctness of good preparation 0-Sims allows us to annihilate i-correctors together with preparation 0-Sims, thereby obtaining the ideal preparation 0-Gas alone. We can now state our basic lemma.

Lemma 7 (Model Conversion). *Consider the measurement-based simulation of operations in a finite quantum universal set, \mathcal{G} , and let C_{sim} be the maximum number of locations in any 0-Sim. Then, if the measurement-based simulation of a quantum circuit expressed in terms of operations in \mathcal{G} is subject to local noise of strength ε , there is a quantum circuit with operations in \mathcal{G} which produces the same output probability distribution and is subject to local noise with strength*

$$\hat{\varepsilon} \leq C_{\text{sim}} \varepsilon$$

(or $\hat{p} \leq C_{\text{sim}} p$ if noise is stochastic).

The proof of this lemma is essentially the same as the proof of the leakage-reduction lemma and will not be discussed in detail. Similar to coherent i-LRUs and their inverses, we can consider the coherent version of an i-corrector which retains the classical record of Pauli corrections. We can insert pairs of coherent i-correctors and their inverses preceding every bad 0-Sim in order to convert them to the simulation of some faulty operations (which may be correlated since they depend on the classical record of the Pauli corrections). Overall, by creating i-correctors out of 0-Sims simulating measurements, moving them to the left of 0-Sims simulating quantum gates, and annihilating them inside 0-Sims simulating qubit preparations, we can convert any noisy measurement-based simulation of an ideal quantum circuit to a direct noisy implementation of the quantum circuit. Then, local

noise afflicting the measurement-based simulation will be equivalent to local noise acting on the equivalent quantum circuit, and the transformation of the noise strength can be estimated as in the proof of lemma 5.

4.3.3 The Threshold Theorem for Quantum Computation by Measurements

The model-conversion lemma tells us that noisy measurement-based quantum computation can be analysed in terms of an equivalent noisy quantum-circuit computation. In particular, if we consider the measurement-based simulation of a fault-tolerant quantum circuit, we expect that this simulation will also be fault tolerant. To prove the existence of an accuracy threshold for such a noisy measurement-based simulation, we can first use the model-conversion lemma to convert it to an equivalent noisy quantum circuit computation that implements the fault-tolerant quantum computation. Then, we can apply theorem 5 to complete the proof. We can thus prove the following theorem.

Theorem 7 (Accuracy Threshold for Measurement-Based Quantum Computation). *Suppose gadgets can be constructed such that they satisfy properties 1 and 2 for a distance- d quantum error-correcting code, and let $t = \lfloor \frac{d-1}{2} \rfloor$. Let c be the largest size and d be the largest depth in any 1-Rec, and let C be the size of the largest 1-exRec. Consider the measurement-based simulation of all operations in our finite quantum universal set, \mathcal{G} , and let C_{sim} be the maximum number of locations in any 0-Sim. Suppose that elementary operations are afflicted by local noise of strength*

$$\varepsilon < \varepsilon_{\text{thr}} \equiv (A')^{-1/t} / C_{\text{sim}} ,$$

(or $p < p_{\text{thr}} \equiv (A')^{-1/t} / C_{\text{sim}}$ if noise is stochastic) with A' as in equation (3.14) where A is the largest number of malignant sets of $t+1$ faults in any 1-exRec and $B \geq (t+1)e^{(C-t-2)\varepsilon_{\text{thr}}} \binom{C}{t+2}$ (or $B = \binom{C}{t+2}$ if noise is stochastic). Then, for any fixed $\delta_0 > 0$, any ideal quantum circuit of size L and depth D can be simulated with error δ_0 by a noisy measurement-based quantum computation of size $O(L(\log L)^{\log_{t+1} c})$ and depth $O(D(\log L)^{\log_{t+1} d})$.

The value of C_{sim} depends on both our universal set and the particular construction of 0-Sims. For example, in the graph-state model, the simulation of the Hadamard gate requires 3 elementary operations (one qubit preparation, one CPHASE gate and one measurement as in figure 4.5), the simulation of the CNOT gate requires $2 \times 3 + 1 = 7$ elementary operations (two Hadamard simulations and one CPHASE gate) and the simulation of the T gate requires $2 \times 3 = 6$ elementary operations (the three operations in figure 4.5 and one Hadamard simulation). Hence, if $\mathcal{G} = \{\text{CNOT}, H, T\}$, then $C_{\text{sim}} = 7$ and the accuracy threshold for measurement-based quantum computation is at most an

order of magnitude lower than for quantum-circuit computation. This conclusion is however overly pessimistic since the construction of 0-Sims can be optimized by carefully choosing our gate set; e.g., [80] gives Monte-Carlo estimates of the accuracy threshold for measurement-based quantum computation which are very close to similar threshold estimates for quantum-circuit computation.

4.4 History and Acknowledgements

A discussion about leakage noise was informally given first in [52] and [67]. However, these discussions were not rigorous since physical leakage noise cannot be accurately captured by the stochastic noise models considered in all proofs of the quantum threshold theorem prior to the results in [49] and [1]. In fact, Terhal started thinking about proving the quantum threshold theorem for leakage noise shortly after the publication of [49], but she found it hard to adapt the proof in [55] to a setting where leakage faults occur. By using the new proof in [1] which was much more transparent, Terhal and I analyzed the problem of fault-tolerant quantum computation subject to local leakage noise in [2]. Later, and with the help of the idea of level reduction, I simplified the proof via the leakage-reduction lemma that was given here.

The problem of fault-tolerant measurement-based quantum computation was first analyzed by Raussendorf in his thesis [81] which discussed cluster-state computation afflicted by a certain type of stochastic noise. A proof of a threshold theorem for measurement-based quantum computation for a more general noise model was later given by Nielsen and Dawson [82]; this proof was based on the proof in [49] for quantum-circuit computation that applies to coherent noise. In fact, Nielsen and Dawson emphasized in [82] that noisy measurement-based simulations of quantum circuits induce effective noise correlations between the simulated quantum gates so that even if physical noise is independent and stochastic, a quantum threshold theorem that applies to coherent faults is necessary. Later, Leung and I gave a simpler proof [83] by showing that noisy measurement-based quantum computation can be viewed as an equivalent noisy quantum-circuit computation where the quantum state is encoded in a random basis that corresponds to the Pauli correction operators—it is correlations due to this coding that Nielsen and Dawson were observing in [82], but the argument in [83] shows that such correlations are no different than the correlations that are induced due to coding in an error-correcting code as explained in §3.3.1. The proof given here is similar to the proof in [83], but it is phrased syntactically via the notion of correctness that is analogous to the semantic notion of *composable simulations* [75] used in [83].

Chapter 5

Lower Bounds on the Quantum Accuracy Threshold

5.1 Introduction

In this chapter, I will present explicit gadgets satisfying properties 1 and 2 and I will use them to establish lower bounds on the quantum accuracy threshold. First, in §5.2, I will describe general gadget constructions that apply to any stabilizer code. Then, in §5.3, I will present specific gadgets for fault-tolerant quantum computation based on the Bacon-Shor code.

In the gadget constructions that I will present, some additional assumptions to those in proposition 2 will be made. Although these assumptions are not essential for proving the quantum threshold theorem, they simplify the presentation and improve our accuracy threshold lower bounds. These assumptions are listed in the following proposition.

Proposition 3. *For our gadget constructions and for deriving our lower bounds on the quantum accuracy threshold, we will make the following additional assumptions to those in proposition 2:*

1. *We can perform quantum measurements during the computation and apply quantum gates conditioned on the measurement outcomes. Furthermore, the speed of quantum measurement is of the same order as the speed of quantum gates.*
2. *The classical processing of measurement outcomes can be done instantaneously and flawlessly.*
3. *Multi-qubit quantum gates can be applied on any subset of qubits irrespective of their physical separation.*
4. *There are no leakage faults.*

As shown in [84], the assumptions about the speed of measurement and the speed of classical processing can be lifted without a significant effect on our accuracy threshold lower bounds. If classical processing is not flawless then, as noted in §3.2, the quantum threshold theorem can still be

proved but the overall accuracy cannot be made arbitrarily small; instead, it will be limited by the accuracy of the gates implementing the classical processing. The quantum threshold theorem can also be proved if geometric constraints are imposed so that multi-qubit gates can only be applied on qubits that are geometrically local [85]; for a two-dimensional square lattice architecture with only nearest-neighbor gates, [86] shows that the accuracy threshold is decreased by a small factor (certainly less than an order of magnitude) compared to the case where no geometric constraints are imposed. Finally, as discussed in §4.2, leakage faults can be dealt with by inserting leakage-reduction units in the quantum circuits preceding every gate. The accuracy threshold in this case decreases by at most an order to magnitude compared to the case when there is no leakage; however, using contracted 0-Recs, it is expected that the accuracy thresholds with and without leakage faults can be brought to be closer.

5.2 Gadget Constructions

5.2.1 Fault-Tolerant Quantum Error Correction

Let us first discuss how to construct 1-EC gadgets satisfying properties 1. The key property guiding the construction is property 1(b) for $s = 0$ and $r \leq t$; i.e., we must assure that when a 1-EC with at most t faults has an input code block that is in the code space, it produces the same output after ideal decoding as if there were no faults inside the 1-EC.

Our 1-EC gadgets will consist of a syndrome measurement procedure which will be possibly followed by the application of some recovery operator. In fact, if the subsequent gates belong in the Clifford group, the recovery operators need not be explicitly applied but they can be kept in a classical memory similar to Pauli corrections in measurement-based quantum computation. Hence, our goal is to construct a quantum circuit that performs syndrome measurement and satisfies properties 1. We will restrict our discussion to stabilizer codes as they are well-suited for fault-tolerant quantum computation; then the code generators are tensor products of Pauli operators.

5.2.1.1 Syndrome Measurement with Cat States

We first note that the code generators commute, so it is sufficient to measure each one separately in order to obtain the syndrome. A method for performing this measurement, which is due to Shor [51], is based on the circuit in figure 5.1 that measures a unitary operator U with eigenvalues ± 1 —since the eigenvalues of the code generators are ± 1 , they can be measured using this circuit. For instance, consider a code generator, G_a , of weight w ; if we view a as a length- $2n$ binary vector,

$$G_a = \bigotimes_{j=1}^n \left(i^{a[j] \cdot a[j+n]} X^{a[j]} Z^{a[j+n]} \right). \quad (5.1)$$

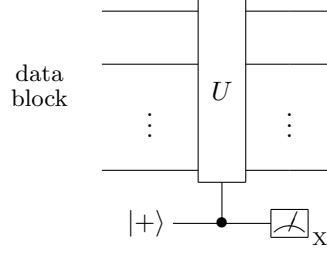


Figure 5.1. A measurement of U using one ancillary qubit. If the measurement outcome is $+1$ (respectively, -1), the data qubits are projected in the $+1$ (respectively, -1) eigenspace of U . The controlled- U gate applies U on the data block if the state of the ancillary qubit is $|1\rangle$, and applies the identity if the state of the ancillary qubit is $|0\rangle$.

Then, the circuit for measuring G_a will consist of an ancillary qubit, a sequence of w two-qubit gates (we apply w controlled- $(i^{a[j] \cdot a[j+n]} X^{a[j]} Z^{a[j+n]})$ gates with control the ancillary qubit and target qubit j in the data block), and a final measurement of the ancillary qubit.

This circuit does *not* yet satisfy properties 1 because the ancillary qubit interacts with multiple qubits in the data block. This is a problem because a single fault acting on the ancillary qubit can give rise to an error that propagates to cause errors on *several* qubits in the data block; e.g., if the error acting on the ancillary qubit is X , then it propagates to become an X error on any qubits in the data block that interact with this ancillary qubit via CNOT gates subsequent to the fault.¹ To prevent such propagation of errors to multiple qubits in the code block, we can replace the one ancillary qubit in figure 5.1 by multiple ancillary qubits. In particular, we can prepare w ancillary qubits in the *cat state*, $|+\rangle_{\text{rep}} \propto |0\rangle^{\otimes w} + |1\rangle^{\otimes w}$, i.e., the logical $|+\rangle$ state of the w -qubit quantum repetition code [51, 87]. Now, the interaction of ancillary and data qubits is bitwise or *transversal*: the j th ancillary qubit interacts via a controlled- $(i^{a[j] \cdot a[j+n]} X^{a[j]} Z^{a[j+n]})$ gate with the j th qubit in the data block. The benefit of transversal interactions is that faults acting on, say, r qubits in the cat state can cause errors on at most r qubits in the data block, thus solving the problem of propagation of errors. The measurement in figure 5.1 needs now to be replaced by a measurement on all w ancillary qubits. This measurement can also be performed *transversally* by separately measuring each of the w ancillary qubits along the eigenbasis of X ; then we can calculate the eigenvalue of G_a by computing the parity of the transversal measurement outcomes. However, in general this eigenvalue cannot immediately be trusted: Since even a single fault might flip one of the transversal measurements causing an error in the computation of the parity of the outcomes, each code generator must be measured repeatedly sufficiently many times depending on the distance of the code; the eigenvalue of G_a will then be obtained by taking a majority vote.

¹It is important to note how X and Z errors are propagated by CNOT gates. A CNOT propagates an X error in its control qubit to X errors in both control and target qubits; it also propagates a Z error in its target qubit to Z errors in both control and target qubits.

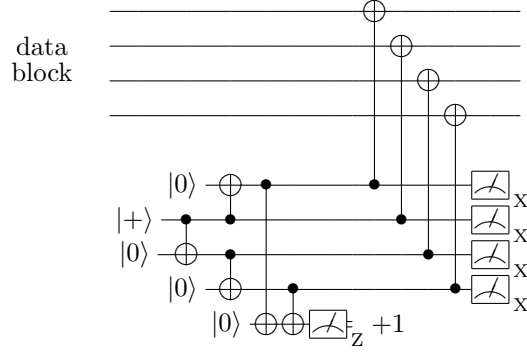


Figure 5.2. Example of syndrome measurement with cat states. In this case, the code generator $X^{\otimes 4}$ is measured. A four-qubit cat state is prepared and verified, and then it controls the application of the Pauli operator to the data block. Finally, all ancillary qubits are measured along the eigenbasis of X , and the parity of the measurement outcomes equals the measured eigenvalue of $X^{\otimes 4}$.

To finish the construction of the 1-EC gadget, it remains to specify how to prepare the cat state. Constructing an encoding circuit for $|+\rangle_{\text{rep}}$ is straightforward: e.g., we may start with the state $|+\rangle \otimes |0\rangle^{\otimes(w-1)}$ and perform CNOT gates with control the first qubit and target every other qubit. To deal with the problem of the propagation of errors inside the encoding circuit, we may next *verify* the cat state [51]. Verification can be performed by preparing extra ancillary qubits and using them to measure operators in the stabilizer of the cat state in order to check for undesired fault patterns inside the encoding circuit—of concern are multiple X errors since any pair of Z errors acts trivially on $|+\rangle_{\text{rep}}$. For instance, figure 5.2 shows the complete circuit for measuring the weight-4 operator $X^{\otimes 4}$. The verification step is a measurement of the operator $Z \otimes I \otimes I \otimes Z$ which belongs in the cat-state stabilizer. If the outcome of the verification measurement is -1 , then a fault might have occurred in one of the two later CNOT gates during cat-state encoding which might have caused X errors on two qubits of the cat state; in this case, the cat state is *discarded* and the encoding is repeated anew. If the outcome of the verification measurement is $+1$, at least two faults are required to introduce two X errors in the cat state, which can therefore be accepted. It is straightforward to adapt the principles underlying the construction of this circuit to measure code generators of higher weight for any stabilizer code.

Due to the transversal interaction between the verified cat state and the data block, property 1(b) is satisfied: Indeed, if the input data block passes through an s -filter, then its state can be expanded as a sum of terms where, in each term, Pauli errors on at most s qubits act on some state in the code space. Also, the r faults inside the 1-EC gadget cannot cause errors acting on more than r qubits in the data block output from 1-EC. Overall, the state of the data block that is output from 1-EC can be expanded as a sum of terms where, in each term, Pauli errors on at most $s + r$ qubits act on some state in the code space. Since $s + r \leq t$, all these Pauli errors can be corrected, which implies that their linear sum will also be corrected. In §5.2.1.4, we will discuss why property

1(a) is also satisfied. Finally, let us note that cat-state verification is *not* necessary for obtaining a 1-EC gadget satisfying properties 1; [84] discusses a more efficient procedure in which verification is avoided and, instead, the cat state is decoded *after* interacting with the data.

5.2.1.2 Syndrome Measurement with Encoded Pauli Eigenstates

A more efficient 1-EC gadget can be constructed for CSS codes [43, 44] for which, as discussed in §1.3.3, each code generator can be chosen to be either X -type or Z -type. This 1-EC construction, which is due to Steane [88], is based on the circuits

$$\begin{array}{c} |\psi\rangle_L \text{---} \oplus \text{---} \\ |0\rangle_L \text{---} \bullet \text{---} \end{array} \begin{array}{c} \diagup \\ \diagdown \end{array} \quad ; \quad \begin{array}{c} |\psi\rangle_L \text{---} \bullet \text{---} \\ |+\rangle_L \text{---} \oplus \text{---} \end{array} \begin{array}{c} \diagup \\ \diagdown \end{array} \quad \begin{array}{c} X \\ Z \end{array},$$

where $|\psi\rangle_L$ is the state of the data block and $|0\rangle_L$, $|+\rangle_L$ denote ancillary blocks encoded in the corresponding logical states in the *same* code as the data. For CSS codes with $k = 1$ logical qubit, the logical CNOT gate can be implemented *transversally*, i.e., by performing bitwise a CNOT from each qubit in the control block to each qubit in the corresponding position in the target block [51, 89]. The CNOT gates in the figure above denote such transversal CNOT gates acting between the data and ancillary blocks, and the measurements are understood as also being performed transversally, i.e., separately on each qubit in the ancillary block.

Evidently, as we would expect for error-correction circuitry, these circuits act trivially on the logical state, $|\psi\rangle_L$. However, their non-trivial content is that the outcomes of the transversal measurements reveal the *parity* of the code syndromes in the data and the ancillary blocks. Indeed, the transversal CNOT gates propagate any Pauli errors in the data block to the two ancillary blocks. Then, the X -type syndrome can be extracted by applying a classical parity check matrix to the measurement outcomes of those ancillary qubits measured along the eigenbasis of X . Specifically, for the X -type generator $X_a \equiv \otimes_{j=1}^n X^{a[j]}$, if the outcomes of the transversal measurements of X are $x = (x_1, x_2, \dots, x_n)$, then the measured eigenvalue of X_a is $a \cdot x \pmod{2}$. The Z -type syndrome can be extracted similarly from the transversal measurements of Z .

As for syndrome measurement with cat states, the logical $|0\rangle$ and $|+\rangle$ ancillary states need to be verified after encoding. Again, verification can be performed by using additional ancillary qubits and measuring suitable check operators. As an alternative to verification, a *purification* procedure can be used in which multiple copies of the logical states are encoded independently and, then, they interact pairwise in order to detect errors during encoding. For instance, one way to purify the $|0\rangle_L$ state against X errors is to prepare another *verifier* $|0\rangle_L$ block, and perform a logical CNOT with the first block as control and the verifier block as target. Finally, each qubit in the verifier block is measured in the computation basis, and a parity check is applied to the measurement outcomes (Fig. 5.3). In this case, not just the Z -type stabilizer generators are extracted for the first block, but

also, after performing classical error-correction on the measurement outcomes, the eigenvalue of the logical Z operator (also a Z -type Pauli operator) can be calculated. The first ancillary block is then discarded if either the syndrome is nontrivial or the eigenvalue of Z_L is -1 . Purification against Z errors can be conducted similarly by reversing the direction of the logical CNOT gate and measuring the verifier qubits in the eigenbasis of X . And, such purification schemes can be iterated if necessary by using the output copies from one purification round as the inputs to a following round. The $|+\rangle_L$ ancillary blocks can be purified similarly.

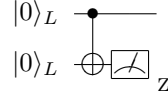


Figure 5.3. A purification procedure against X errors for logical $|0\rangle$ blocks encoded in some CSS code. Two blocks are encoded separately (the encoding circuits are not shown) and, then, they interact via a logical CNOT followed by measuring all qubits in the second block in the computation basis. Classical parity checks on the measurement outcomes allows to detect X errors and also calculate the eigenvalue of the logical Z operator on the first block. If no errors are detected and the eigenvalue of Z_L is $+1$, then the first block is accepted; otherwise it is rejected.

Because of the transversal interaction between the data and ancillary blocks, this 1-EC gadget satisfies property 1(b); we will discuss why property 1(a) is also satisfied in §5.2.1.4. Finally, we note that the purification procedure describe above, like the verification procedure for cat states, is nondeterministic: there are probabilistic fluctuations in the number of ancillary blocks that must be prepared before the first ancillary block is accepted. However, both methods can be adapted to become deterministic by increasing the measured check operators in verification procedures, or the number of verifier blocks in purification procedures. Alternatively, as for cat-state verification, the purification of $|0\rangle_L$ and $|+\rangle_L$ blocks can be avoided altogether and replaced by a suitable decoding step *after* the ancillary blocks interact with the data block [84].

5.2.1.3 Syndrome Measurement via Logical Teleportation

A third method for syndrome measurement that applies to any stabilizer code can be based on quantum teleportation. In this construction, which is due to Knill [59], the 1-EC gadget implements the *logical* teleportation of the state of the input data block while at the same time extracting the code syndrome. Figure 5.4 shows a schematic of Knill’s 1-EC gadget: Two ancillary blocks are prepared in a logical Bell state, $|\Phi_0\rangle_L \propto |0\rangle_L|0\rangle_L + |1\rangle_L|1\rangle_L$, and then the data and one ancillary block interact via *transversal* CNOT gates followed by transversal measurements on all their qubits. Finally, a *logical* Pauli operator is applied on the second ancillary block in order to complete the *logical* teleportation of the state of the data block.

The outcomes of the transversal measurements reveal the parity of the code syndromes in the data and ancillary blocks. First, this syndrome information allows us to perform classical error

correction on the measurement outcomes. Then, we can compute parities on the, now *corrected*, measurement outcomes in order to obtain the eigenvalues of the *logical* $X \otimes X$ and $Z \otimes Z$ operators acting on the first two blocks whose qubits were measured. Finally, by knowing the eigenvalues for these two logical operators, we can complete the *logical* teleportation of the state of the data block to the second ancillary block up to some appropriate logical Pauli correction operator. As in the previous two methods for syndrome measurement, the two ancillary blocks for this method need to be verified or purified before they are used. Overall, after a suitable verification or purification procedure is applied, the 1-EC gadget in figure 5.4 satisfies property 1(b); again, property 1(a) will be discussed separately in §5.2.1.4.

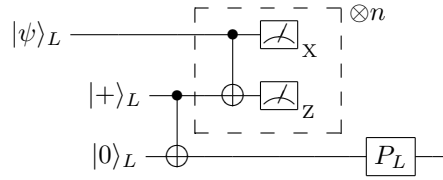


Figure 5.4. A schematic of Knill’s 1-EC gadget for a code with n physical qubits in the block. Two ancillary blocks are prepared in a logical Bell state and, then, transversal CNOT gates and measurements are performed. The measurement outcomes are constrained by parities which enable classical error correction. The *corrected* measurement outcomes allow us to determine the logical Pauli operator, P_L , necessary to complete the *logical teleportation* of the input data state, $|\psi\rangle_L$.

A remarkable feature of this 1-EC gadget is that faults in the transversal CNOT gates and the subsequent transversal measurements—which combined can be viewed as performing transversal *Bell measurements*, i.e., measurements along the two-qubit basis $\{(Z^{j_1} X^{j_2} \otimes I)|\Phi_0\rangle \mid j_1, j_2 = 0, 1\}$ —can only result in an error in the logical Pauli correction operator. In other words, as long as the two ancillary blocks are prepared (and verified) without faults, the output code block is always in the code space independent of whether faults occurred in the transversal Bell measurements. The value of this observation comes when performing an analysis of malignant set of locations inside a 1-exRec where such 1-EC gadgets are used: Consider the leading 1-EC gadgets inside the 1-exRec and let faults occur in the transversal Bell measurements inside them (while more faults may have occurred elsewhere in the 1-exRec). Then, since these faults cannot take the output code block outside the code space, these faulty locations cannot belong to malignant sets of locations for the 1-exRec. Therefore, for the purposes of an analysis of malignant sets of locations, the transversal Bell measurements in the *leading* 1-EC gadgets in all 1-exRecs can be taken to be effectively ideal!

5.2.1.4 Satisfying Property 1(a)

Property 1(a) is the statement that independent of the state of the input block, a 1-EC with at most r faults outputs a block which always passes through an r -filter for $0 \leq r \leq t$. In particular,

applying the property for $r = 0$ requires that a 1-EC gadget without faults takes any input block to an output block which is in the code space. If each possible value of the syndrome corresponds to a correctable error (in which case the code is called *perfect*), then clearly the output of a faultless 1-EC will be in the code space after applying the necessary recovery operator. But, even if the code is not perfect, we may always assign to every syndrome that does not correspond to a correctable error some Pauli operator (chosen by convention) that returns the subspace labeled by the syndrome to the code space (see the last comment in §1.3.2).

At this point, it is convenient to include some reference system, R , that *purifies* the state of the input data block. This state can then be written as a sum of terms where, for each term, the state of the code block is a vector, $E_a|\psi_a\rangle_L$, and the state of the reference system is $|a\rangle_R$. Here $\{E_a\}$ are Pauli operators that label the orthogonal subspaces corresponding to all possible syndromes, $\{|\psi_a\rangle_L\}$ are some states in the code space (that may depend on a), and $\{|a\rangle_R\}$ need not be normalized nor mutually orthogonal. In this language, a faultless 1-EC realizes an isometry that acts on the data block and an ancillary system, A , that is initialized in the state $|e_0\rangle_A$ as

$$\sum_a E_a|\psi_a\rangle_L \otimes |a\rangle_R \otimes |e_0\rangle_A \xrightarrow{\text{EC}}^{\text{faultless}} \sum_a \mathcal{O}_a|\psi_a\rangle_L \otimes |a\rangle_R \otimes |e_a\rangle_A, \quad (5.2)$$

where $\{\mathcal{O}_a\}$ are logical operators, i.e., they have support only on the code space, and $\{|e_a\rangle_A\}$ are also not assumed to be normalized nor mutually orthogonal. Thus, as desired, after tracing over the reference and ancillary systems, a faultless 1-EC gadget produces an output which is, in general, some density matrix in the code space.

We can now show property 1(a) is satisfied even if $0 < r \leq t$ faults occur inside the 1-EC gadget. By construction, in all 1-EC gadgets we have considered, the data and ancillary blocks interact via transversal gates. Moreover, the preparation and verification (or purification) of the ancillary blocks is done in a way that prevents $0 < r' \leq t$ faults in this part of the 1-EC gadget from acting nontrivially on more than r' qubits in an accepted ancillary block. Overall, each of the r faults inside a 1-EC gadget can only act nontrivially on the *same* r qubit positions in the data and ancillary blocks.

Let \mathcal{I}_r denote this set of at most r qubit positions in the data or ancillary blocks that may have errors due to the r faults inside the 1-EC gadget. Consider one term in the expansion of the state of the input data block, $\sum_a E_a|\psi\rangle_L \otimes |a\rangle_R$, where the input data block has Pauli error E_a . Also, expand the r fault operators in terms of Pauli operators and pick one term in this expansion where the Pauli operator $E_b^{(\mathcal{I}_r)}$ acts on the data block and $E_c^{(\mathcal{I}_r)}$ acts on the ancillary blocks. (Here, our notation is that $E_*^{(\mathcal{I}_r)}$ is a tensor product of Pauli operators with support on the qubits in \mathcal{I}_r alone.) For this particular term, the syndrome will correspond to the error operator $E_c^{(\mathcal{I}_r)}E_a$, so that $E_{\text{rec}}^\dagger E_c^{(\mathcal{I}_r)}E_a \equiv \mathcal{L}$ where E_{rec}^\dagger is the recovery operator and \mathcal{L} is some logical operator. Then,

after the recovery, E_{rec}^\dagger , is applied to the data block that is in the state $E_b^{(\mathcal{I}_r)} E_a |\psi\rangle_L \otimes |a\rangle_R$, the combined operator that acts on $|\psi\rangle_L \otimes |a\rangle_R$ is

$$E_{\text{rec}}^\dagger E_b^{(\mathcal{I}_r)} E_a = \mathcal{L} E_a^\dagger \left(E_c^{(\mathcal{I}_r)} \right)^\dagger E_b^{(\mathcal{I}_r)} E_a = \mathcal{L} E_d^{(\mathcal{I}_r)}, \quad (5.3)$$

where the crucial observation is that $E_d^{(\mathcal{I}_r)}$ is only supported on the qubits in \mathcal{I}_r . We conclude that, for each term in our expansion, the output of an r -good 1-EC deviates from the code space by the action of Pauli operators acting on at most r qubits in the code block. Hence, since every term in our expansion passes through an r -filter, by linearity, the linear sum of all terms will also pass through an r -filter.

This completes our general discussion of how 1-EC gadgets satisfying properties 1 can be constructed.

5.2.2 Encoded Quantum Universality

Let us now discuss how to construct a universal set of 1-Ga gadgets satisfying properties 2. This construction is guided by the C_k gate hierarchy in definition 2. We have already discussed that the Clifford group, C_2 , is *not* dense in $SU(2^n)$. In fact, the Gottesman-Knill theorem [90] shows that quantum computation using qubits initialized in eigenstates of Pauli operators, Clifford-group gates, and measurements of Pauli operators can be simulated efficiently with a classical computer—let us call this set of operations *stabilizer operations* or $\mathcal{G}_{\text{stab}}$. But, by adding any *one* non-Clifford gate to $\mathcal{G}_{\text{stab}}$ is sufficient to give a new set of operations which *is* dense in $SU(2^n)$ [33] and, hence, is quantum universal; examples of such gate sets were given in proposition 1.

5.2.2.1 Encoded Clifford-Group Operations

Although 1-Ga gadget constructions are known that apply to any stabilizer code [89], of practical interest are codes for which these constructions are efficient and give high accuracy thresholds. In fact, for some codes the construction of 1-Gas for the gates $\{H, S, \text{CNOT}\}$ that generate the Clifford group is particularly simple. Let us call a 1-Ga *transversal* if it can be realized by 0-Gas that act bitwise on the qubits in the code block (for a 1-Ga simulating a single-qubit gate), or bitwise between qubits at corresponding positions across different code blocks (for a 1-Ga simulating a multi-qubit gate). By construction, it is clear that transversal 1-Gas satisfy properties 2.

CSS codes provide a family of codes with such useful transversality properties [51, 89]: For any CSS code with $k = 1$ logical qubit, the logical CNOT can be implemented by bitwise CNOT gates. Moreover, if the CSS code is constructed from a self-dual classical code (i.e., if the quantum check matrix in equation (1.32) can be written so that $H = H'$), then the logical Hadamard gate can be implemented by bitwise Hadamard gates. If, in addition, this self-dual classical code is doubly even

(i.e., if all its code words have Hamming weight multiple of 4), then the logical S gate is transversal so that the logical Clifford group can be generated by 1-Gas which are transversal. An example of a code with these properties is Steane's $[[7, 1, 3]]$ code [37] discussed in §1.3.4.

5.2.2.2 Quantum Software and Encoded Non-Clifford Gates

To complete our construction of 1-Gas satisfying properties 2 for a universal set of gates, it remains to construct a 1-Ga for a non-Clifford gate. This construction can be based on preparing a suitable ancillary state and then using a gate teleportation circuit [51, 42, 32, 77].

Consider the circuit shown in figure 5.5: To realize the single-qubit rotation by angle θ around the z -axis, $U_z(\theta) = \exp(-i\frac{\theta}{2}Z)$, we first prepare an ancillary qubit in the state $|A_\theta\rangle \equiv U_z(\theta)|+\rangle$ and, then, we perform a CNOT gate with the data qubit as control and the ancillary qubit as target. Finally, we measure the ancillary qubit in the computation basis. If the $+1$ eigenvalue is obtained, then we have successfully teleported the gate $U_z(\theta)$; otherwise, when the outcome is -1 , $U_z(-\theta)$ has been teleported instead and the correction operator $U_z(2\theta)$ needs to be applied.

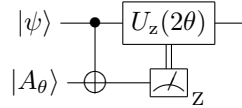


Figure 5.5. A circuit that *teleports* the gate $U_z(\theta)$. An ancillary qubit is prepared in the state $|A_\theta\rangle \equiv U_z(\theta)|+\rangle$, and then a CNOT gate is executed as shown. Finally, the ancillary qubit is measured in the computation basis and the correction gate $U_z(2\theta)$ is applied if the measurement outcome is -1 .

We observe that for $\theta = \pi/4$, the teleported gate is $T \equiv U_z(\frac{\pi}{4}) \notin C_2$. Furthermore, in this case the correction gate is $S \equiv U_z(\frac{\pi}{2}) \in C_2$, and, so, a 1-Ga for simulating it can easily be constructed. It remains to construct a procedure for preparing the *logical* $|A_{\pi/4}\rangle$ state such that the 1-Ga for the T gate satisfies properties 2. Since the logical $|A_{\pi/4}\rangle$ is just an ancillary state, its preparation can be done separate to the main computation; we can then view this quantum state as a form of “software” that we prepare and consume in order to implement fault-tolerant quantum computation.

The logical $|A_{\pi/4}\rangle$ state is the $+1$ eigenstate of the *logical* Clifford-group operator $TXT^\dagger = T^2X = SX$. Since we can construct a 1-Ga for the gate SX , we can also measure this operator using cat states: That is, we can measure the logical SX by preparing a cat state, controlling each gate in the 1-Ga for SX from a different cat-state qubit and, finally, measuring all qubits of the cat state along the eigenbasis of X and computing the parity of the measurement outcomes. Since a single fault in this measurement can cause an error in the eigenvalue of the logical SX operator that is measured, the measurement needs to be repeated sufficiently many times using different cat states with error-correction steps inserted between every two successive measurements.

With this procedure which applies to any stabilizer code, we can therefore construct a 1-Ga for

the non-Clifford gate T that satisfies properties 2. An alternative construction for the non-Clifford Toffoli gate will be given in §5.3.4. Together with the 1-Ga constructions for Clifford-group gates discussed in the previous section, we have constructed 1-Gas satisfying properties 2 for a quantum universal set of gates.

5.2.3 Fault-Tolerant Preparation and Measurement

Finally, let us briefly discuss how 1-Ga gadgets for preparation of Pauli eigenstates and measurements in the eigenbases of Pauli operators can be constructed. As with Clifford-group gate 1-Ga constructions, efficient preparation and measurement 1-Gas can be constructed in the special case of CSS codes. Indeed, for CSS codes, the 1-Ga for the measurement of Z consists of measurements of all qubits along the eigenbasis of Z , followed by classical postprocessing of the measurement outcomes. This postprocessing consists of classical error correction on the measurement outcomes (here, it is important that for CSS codes the stabilizer generators are only X - and Z -type), followed by computing the parity of the corrected outcomes to obtain the eigenvalue of the logical Z . A similar procedure can be used to construct a 1-Ga for measuring the operator X . Also, for CSS codes, there are efficient preparation 1-Gas for Pauli eigenstates. The construction of these 1-Gas can be based on a *purification* procedure in which multiple copies of the same logical state are encoded separately, and then they interact pairwise in order to detect errors similar to figure 5.3.

For general stabilizer codes, preparation and measurement 1-Gas are more complex. First, 1-Gas for measuring Pauli operators can be constructed by repeated measurements of the corresponding logical Pauli operator using cat states. Then, the majority of the different measurement outcomes gives the eigenvalue of the Pauli operator that is measured by the 1-Ga. Similarly, 1-Gas for preparing Pauli eigenstates implement measurements of the code generators in order to transform any arbitrary initial state to a state in the code space, followed by a sufficient number of repeated measurements of the logical Pauli operator whose eigenstate we want to prepare. (For both preparation and measurement 1-Gas, error-correction steps need to be inserted between the repeated measurements in order to satisfy properties 2.)

This completes our discussion of constructing preparation and measurement 1-Gas. At this point, we have shown how to obtain 1-Gas satisfying properties 2 for a universal set of operations (preparation of Pauli eigenstates, a quantum universal set of gates, and measurement of Pauli operators).

5.3 Quantum Fault Tolerance with the Bacon-Shor Code

At last, it has come time to establish a numerical lower bound on the value of the quantum accuracy threshold. For simplicity, in the remaining discussion we will only consider *stochastic* local noise.

Similar techniques can be used to give lower bounds of the same order for local noise that is not stochastic; however, the calculations are easier to perform and present in terms of probabilities instead of quantum amplitudes.

The fault-tolerant quantum circuits we will consider will be encoded using a subsystem code described by Bacon in [91]—because of the close relation of this code with Shor’s code [36], we will refer to it as the *Bacon-Shor code*. There is a different Bacon-Shor code for every integer $n > 1$; for fixed n , the corresponding code, $\mathcal{C}_{BS}^{(n)}$, is a distance- n stabilizer CSS code encoding one protected logical qubit into a code block of n^2 physical qubits.

It is convenient to imagine placing the n^2 qubits in the $\mathcal{C}_{BS}^{(n)}$ block on the *vertices* of an $n \times n$ square lattice (Fig. 5.6). Then, the code’s stabilizer group is

$$\mathcal{S} = \langle X_{j,*} X_{j+1,*} ; Z_{*,j} Z_{*,j+1} \mid j \in \mathbb{Z}_{n-1} \rangle, \quad (5.4)$$

where $O_{j,*}$, $O_{*,j}$ denote operators that act nontrivially as a tensor product of O operators on all qubits in row or column j , respectively. Throughout this section, \mathbb{Z}_m is understood to indicate the set $\{1, 2, \dots, m\}$.

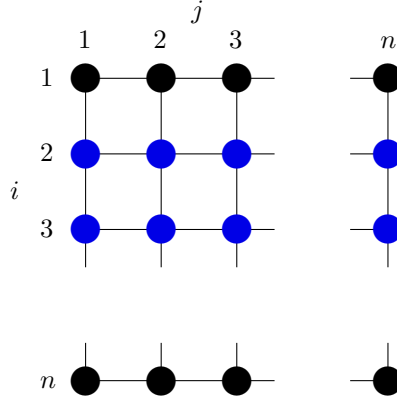


Figure 5.6. Qubits in the $\mathcal{C}_{BS}^{(n)}$ block are placed on the vertices of an $n \times n$ square lattice. An element of the code stabilizer, the operator $X_{2,*} X_{3,*}$ applies X on all qubits shown in blue.

The code *syndrome*, e , induces a decomposition of the Hilbert space, \mathcal{H} , of the n^2 qubits in the code block into subspaces encoding $n^2 - 2(n-1) = (n-1)^2 + 1$ logical qubits. Therefore, within each subspace with fixed syndrome—and, in particular, within the *code space* corresponding to the trivial syndrome—we can define a subsystem decomposition

$$\mathcal{H} = \bigoplus_e (\mathcal{H}_L \otimes \mathcal{H}_T) , \quad (5.5)$$

where we associate \mathcal{H}_L with the one logical qubit protected by the full distance, n , of the code. The logical Pauli operators for this logical qubit can be defined as $X_L = X_{1,*}$ (i.e., a tensor product of X

operators applied on all qubits in the first row) and $Z_L = Z_{*,1}$ (i.e., a tensor product of Z operators applied on all qubits in the first column). The remaining $(n-1)^2$ logical qubits live in \mathcal{H}_T and their logical Pauli operators can be chosen from the non-abelian group

$$T = \langle X_{j,i} X_{j+1,i} ; Z_{i,j} Z_{i,j+1} \mid i \in \mathbb{Z}_n ; j \in \mathbb{Z}_{n-1} \rangle , \quad (5.6)$$

where $O_{i,j}$ denotes an operator O acting on the qubit with coordinates (i, j) . Indeed, the operators in T commute with every operator in the code stabilizer, they commute with the logical operators X_L and Z_L and, furthermore, they can be grouped into $(n-1)^2$ independent pairs of anticommuting operators with operators in different pairs commuting.

Given some nontrivial syndrome value, error recovery for the logical qubit encoded in \mathcal{H}_L proceeds in a similar manner as in the classical repetition code: The eigenvalues of the stabilizer generators $\{X_{j,*} X_{j+1,*}\}$ can be used to correct Z errors on up to $\lfloor n/2 \rfloor$ rows. Moreover, only the parity of Z errors in each row is relevant: an operator acting as Z on a pair of qubits at the same row is an operator in T and, therefore, has no effect on the protected information in \mathcal{H}_L . Error recovery for X errors proceeds similarly along the columns. Further intuition on the Bacon-Shor code can be obtained by considering its derivation from other subspace codes; see chapter A in the appendix.

Since the logical Pauli operators for the logical qubits encoded in \mathcal{H}_T act nontrivially on only two qubits in the code block, if we take into consideration all $(n-1)^2 + 1$ logical qubits then $\mathcal{C}_{BS}^{(n)}$ has distance 2 and it is an *error-detecting* code. However, if we only consider the logical qubit encoded in \mathcal{H}_L , we effectively obtain a distance- n code and errors with support on up to $\lfloor n/2 \rfloor$ qubits in the code block can be corrected—we will call this logical qubit the *protected* qubit. In fact, error recovery for the protected qubit may unavoidably result in applying at the same time nontrivial logical operations with support in \mathcal{H}_T . This is not a problem as long as we never encode useful information in \mathcal{H}_T . We can think of the $(n-1)^2$ logical qubits encoded in \mathcal{H}_T as *gauge* qubits since they correspond to degrees of freedom for the logical information encoded in \mathcal{H}_L . In some cases it will be sufficient to completely disregard the state of the gauge qubits. More interestingly, we will next discuss how fault-tolerant error correction for the protected qubit can be simplified by taking advantage of the presence of gauge qubits.

5.3.1 Syndrome Measurement using the Gauge Qubits

Let us first explain how we can extract the code syndrome indirectly by manipulating the state of the gauge qubits. Our first observation is that we can express the stabilizer generators as

$$X_{j,*} X_{j+1,*} = \bigotimes_{k=1}^n (X_{j,k} X_{j+1,k}) ; \text{ and } Z_{*,j} Z_{*,j+1} = \bigotimes_{k=1}^n (Z_{k,j} Z_{k,j+1}) . \quad (5.7)$$

What is remarkable about this decomposition is that the operators in parentheses are supported on \mathcal{H}_T ; hence, they commute with *all* stabilizer generators and also commute with the logical operators for the protected qubit. Because of this, we can *measure* each of them *separately* and then equation (5.7) implies that the code syndrome can be computed by taking the appropriate parities of the measurement outcomes. Moreover, since these operators act nontrivially on only two qubits in the code block, measuring each one of them is especially easy; e.g., figure 5.7 shows simple circuits for measuring $X_{j,k}X_{j+1,k}$ and $Z_{k,j}Z_{k,j+1}$.

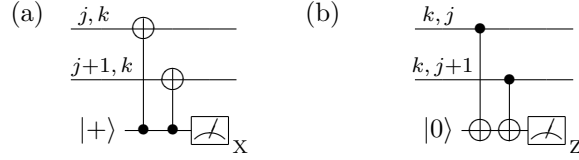


Figure 5.7. (a) A circuit for measuring the operator $X_{j,k}X_{j+1,k}$ using one ancillary qubit. (b) A similar circuit for measuring $Z_{k,j}Z_{k,j+1}$.

Let us now check that these circuits can be used to construct a 1-EC gadget that satisfies properties 1. Consider the circuit in figure 5.7(a); similar observations apply to the other circuit. A fault in the preparation of the $|+\rangle$ state can only cause a Z error that will propagate to flip the measurement outcome but which cannot harm the data. Consider now faults in the CNOT gates. We might worry that there is a problem if the first CNOT is faulty because an X error can result in both data qubits. But we soon realize that $X_{j,k}X_{j+1,k}$ is not an error—it is exactly the operator we are trying to measure and so acts trivially after the measurement has taken place. Finally, we note that a single fault may cause one X and one Z error in *different* qubits in the data block which, therefore, will not pass through a 1-filter. However, as we will discuss in §5.3.3, the only extended rectangles that appear at the lowest levels of the fault-tolerant recursive simulation correspond to the logical CNOT and Hadamard operations. In this case, we may define the 1-filter in these levels as treating X and Z errors separately, and then our 1-EC gadgets satisfy properties 1². Because of this observation, we will henceforth concentrate our discussion on the number of errors of the same “type” (i.e., either X or Z) that faults may produce within the same code block. We should finally note that the measurement outcomes from the circuits in figure 5.7 cannot, in general, be trusted unless some of the measurements are repeated some number of times depending on the distance of the code. This redundancy is necessary since a single fault can cause errors in both the data block and the measurement outcomes. We will return to this point in §5.3.3 by giving specific examples.

The benefit of this syndrome-measurement method is that it significantly reduces the qubit overhead for fault-tolerant error correction: Unlike in the standard 1-EC gadget constructions discussed in §5.2.1, this method does not require preparing and verifying entangled ancillary states—ancillary

²Here it is important that CNOT gates do not mix X and Z errors upon conjugation and Hadamard gates transform X errors to Z errors and vice versa.

qubits initialized in the $|0\rangle$ or $|+\rangle$ state are sufficient as in figure 5.7. For maximum qubit efficiency but at a cost of extra memory error, even a *single* ancillary qubit would suffice to sequentially measure all weight-two gauge-qubit operators necessary to extract the code syndrome. Furthermore, the specific gauge-qubit operators to be measured can be chosen to have support on neighboring physical qubits when qubits in the code block are arranged on a two-dimensional square lattice. For this reason, this method for syndrome measurement may prove to be especially advantageous for geometrically local quantum computing architectures such as, e.g., those envisioned for ion-trap or solid-state implementations.

5.3.2 Standard Syndrome Measurement Procedures

In settings without geometric locality constraints for the interaction of qubits or when qubit movement is much less noisy than quantum gates, syndrome measurement procedures that use encoded ancillary states and transversal interactions between the data and ancillary blocks will give the best accuracy thresholds. Since $C_{BS}^{(n)}$ is a CSS code, Steane's 1-EC gadget presented in §5.2.1.2 can be used to extract the syndrome provided we can prepare and verify logical $|0\rangle$ ($|0\rangle_L$) and logical $|+\rangle$ ($|+\rangle_L$) states for the protected qubit. Alternatively, Knill's 1-EC gadget in §5.2.1.3 requires preparing and verifying a logical Bell state ($|\Phi_0\rangle_L \propto |0\rangle_L|0\rangle_L + |1\rangle_L|1\rangle_L$).

As is evident from the decomposition (5.5), the distinctive feature of subsystem codes is that logical states in \mathcal{H}_L are not uniquely encoded: after having specified a logical state in \mathcal{H}_L , the state in \mathcal{H}_T can still be arbitrary. This freedom in choosing the state of the gauge qubits can be used to our advantage in the design of encoding circuits of logical states for the protected qubit. In particular, we will next discuss how, by exploiting this freedom, we can design remarkably simple encoding circuits for the logical ancillary states required for Steane's and Knill's 1-EC gadgets, thus also reducing the overhead associated with the post-encoding verification of these states.

For concreteness, consider designing an encoding circuit for $|0\rangle_L$, i.e., a state in the code space which is the $+1$ eigenstate of Z_L . With the state in \mathcal{H}_L specified, we can choose the state in \mathcal{H}_T to be the $+1$ eigenstate of the gauge-qubit operators $\{X_{i,j}X_{i+1,j}\}$. In other words, our encoding circuit prepares the $+1$ eigenstate of the operators in the following stabilizer group:

$$\mathcal{S}(|0\rangle_L) = \langle X_{i,j}X_{i+1,j} ; Z_{*,j} \mid i \in \mathbb{Z}_{n-1}; j \in \mathbb{Z}_n \rangle. \quad (5.8)$$

We recognize the state described by equation (5.8) as a tensor product of n *cat states* in the Hadamard-rotated basis, each one lying across a column in figure 5.6. In other words, our $|0\rangle_L$ is the state

$$|0\rangle_L = \bigotimes_{j=1}^n \left(\frac{|++\cdots+\rangle_{\text{col};j} + |--\cdots-\rangle_{\text{col};j}}{\sqrt{2}} \right), \quad (5.9)$$

where inside the parenthesis we have the state of the n qubits in column j .

We can obtain the state $|+\rangle_L$ by applying a logical Hadamard transformation to the state $|0\rangle_L$. We observe that applying Hadamard gates *bitwise* has the same effect as a logical Hadamard gate up to a 90-degree rotation of the square lattice in figure 5.6 (and, also, up to a nontrivial operation acting on the gauge qubits). We can therefore obtain $|+\rangle_L$ by first preparing $|0\rangle_L$, applying Hadamard gates bitwise and, finally, rotating the lattice by 90 degrees. The bitwise Hadamard gates will transform the Hadamard-rotated cat states to become usual cat states in the computation basis. And then, rotating the lattice by 90 degrees will align the n cat states each to lie across a row. Hence, our $|+\rangle_L$ state is

$$|+\rangle_L = \bigotimes_{i=1}^n \left(\frac{|00\cdots 0\rangle_{\text{row:i}} + |11\cdots 1\rangle_{\text{row:i}}}{\sqrt{2}} \right). \quad (5.10)$$

Finally, logical Bell states to be used in Knill's 1-EC gadget can be constructed by interacting two blocks encoded in the states $|+\rangle_L$ and $|0\rangle_L$ via a logical CNOT gate. Since the logical CNOT gate can be implemented by transversal CNOT gates, this construction is also especially simple. Specific examples of Steane's and Knill's 1-EC gadgets will be given in §5.3.3.

5.3.3 Accuracy Threshold Lower Bounds

We can now discuss how to construct 1-Gas for the Bacon-Shor code. As mentioned above, the logical CNOT and Hadamard gates have simple transversal implementations. Let us call *CSS operations* the operations in the set

$$\mathcal{G}_{\text{CSS}} = \{\text{CNOT}, H, \mathcal{P}_{|0\rangle}, \mathcal{P}_{|+\rangle}, \mathcal{M}_X, \mathcal{M}_Z\}, \quad (5.11)$$

where $\mathcal{P}_{|\phi\rangle}$ denotes the preparation of the single-qubit state $|\phi\rangle$ and \mathcal{M}_A denotes the measurement of the single-qubit operator A . Together with the operations in \mathcal{G}_{CSS} , the phase gate S is sufficient to generate the Clifford group and further adding, e.g., the Toffoli gate gives a quantum universal set of operations. To obtain a lower bound on the accuracy threshold for CSS operations, we will perform an analysis of malignant sets of locations in the CNOT 1-exRec. We note that we need not analyse the Hadamard 1-exRec since it contains fewer locations than the CNOT 1-exRec. The same observation is true for the preparation and measurement 1-exRecs.

The reason why the accuracy threshold for CSS operations can be analysed separately is that error correction for the Bacon-Shor code (and, more generally, for all CSS codes) can be implemented using CSS operations alone. Therefore, the extended rectangles corresponding to the logical phase and Toffoli gates need not appear at the lower levels of the recursive fault-tolerant simulation. It follows that if the physical noise strength is below the accuracy threshold for CSS operations, a recursive simulation of CSS operations can achieve any desired accuracy so that, effectively, CSS operations can be viewed as being *nearly ideal* at the highest levels of the recursion. Given highly

reliable CSS operations, our next step will be to show that we can distill highly reliable quantum software (see §5.2.2.2) that will enable the simulation of the logical phase and Toffoli gates that complete quantum universality. Overall, our strategy is to first determine the threshold for CSS operations which, by being significantly lower than the threshold for the distillation of quantum software, will set the accuracy threshold for all operations in our quantum universal operation set.

In the remaining of this section, I will first discuss the accuracy threshold for CSS operations for recursive simulations based on the Bacon-Shor code. Then, in §5.3.4, I will give more details on the simulation of the logical phase and Toffoli gates which complete quantum universality.

5.3.3.1 The 9-Qubit Bacon-Shor Code

The 1-EC Gadgets

Syndrome measurements using the gauge qubits for the 9-qubit Bacon-Shor code, $\mathcal{C}_{BS}^{(3)}$, can be done with the circuits in figure 5.8. For this code, the measurement of one extra operator is necessary in order to satisfy properties 1; in figure 5.8 we have chosen to measure the extra operator $X_{1,j}X_{3,j}$. This additional measurement provides the necessary redundancy in order to avoid the occurrence of more than one error in the data block due to a single fault³.

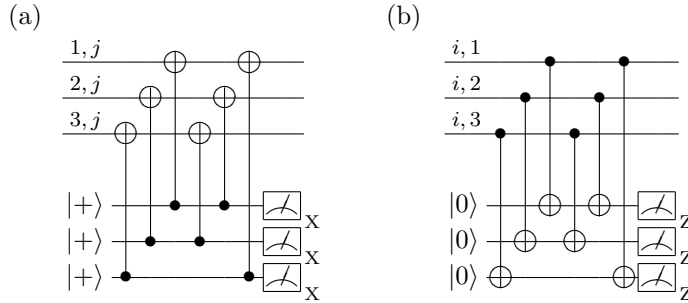


Figure 5.8. A fragment of the 1-EC gadget for $\mathcal{C}_{BS}^{(3)}$ where the syndrome is measured using the gauge qubits. (a) Measuring operators in T along column j . (b) Measuring operators in T along row i . Identical circuits are run in parallel in all rows and columns. In both circuits, the third measurement is used to provide an extra redundancy bit that is needed to satisfy properties 1.

Alternatively, Steane’s 1-EC gadget requires preparing a logical $|0\rangle$ state for extracting the syndrome for Z errors. Since our logical $|0\rangle$ state consists of three Hadamard-rotated cat states aligned across the three columns, the preparation circuit for column j can be as in figure 5.9(a). To obtain the syndrome for X errors, we need to also prepare a logical $|+\rangle$ state which consists of cat states along the three rows; for row i , the preparation circuit is shown in figure 5.9(b). We observe that the

³We are concerned with the case where, e.g., a Z error occurs on both qubits on the support of the second CNOT operating on qubit $(2,j)$ —this error will flip the eigenvalue of the measured operator $X_{1,j}X_{2,j}$, and it will also create a $Z_{2,j}$ error in the data. An error in the eigenvalue of $X_{1,j}X_{2,j}$ will lead us to incorrectly infer an error $Z_{1,j}$ in the data if we do not perform the extra measurement of $X_{1,j}X_{3,j}$. Combining $Z_{1,j}$ with $Z_{2,j}$ leads to an uncorrectable error in the data that is caused by a single faulty CNOT.

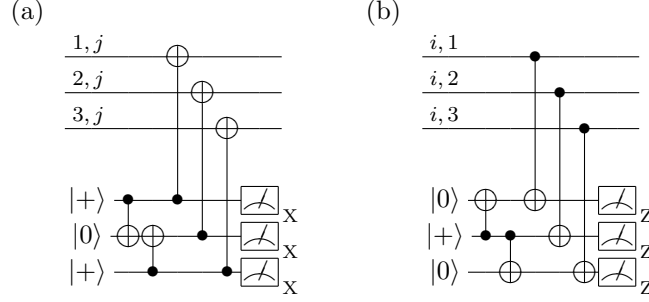


Figure 5.9. A fragment of Steane's 1-EC gadget for $\mathcal{C}_{BS}^{(3)}$. (a) Collecting the syndrome for Z errors. The circuit in column j is shown. (b) Collecting the syndrome for X errors. Here, the circuit for row i is shown. Identical circuits are run in parallel in all rows and columns.

1-EC gadget constructed with these circuits has the remarkable feature that it satisfies properties 1 *without* the need to verify the ancillary states. Indeed, consider a faulty CNOT inside the encoding circuit for these cat states. An X error on both qubits on the support of this gate is either (i) an operator in the cat-state stabilizer or, (ii) is equivalent to a single X error on the other qubit. Similarly for a Z error acting on both qubits. Otherwise, an X error acting on one qubit and a Z error on the other cannot lead to more than one X or one Z error propagating from the cat state to the data block. Knill's 1-EC gadget for this code has the same feature as well; again, properties 1 are satisfied without the need to verify the ancillary logical Bell state which is prepared by encoding three-qubit cat states and then applying transversal CNOT gates.

Accuracy Threshold Lower Bounds

A lower bound on the accuracy threshold can be obtained from theorem 4 by counting the total number of pairs of locations inside the CNOT 1-exRec since it is the largest among all 1-exRecs corresponding to CSS operations. Both Steane's and Knill's 1-EC gadgets contain 72 qubit-preparation, measurement, gate or memory locations. Therefore, the CNOT 1-exRec contains $4 \times 72 + 9 = 297$ locations, which implies $p_{\text{thr}} \geq \binom{297}{2}^{-1} \geq 2.2 \times 10^{-5}$.

We can refine this lower bound by using the notion of malignant sets of locations described in §3.4.1. Let us denote by p_i the noise strength for location type i , where locations are labeled as (1) memory, (2) $|0\rangle$ preparation, (3) $|+\rangle$ preparation, (4) measurement of X , (5) measurement of Z and (6) CNOT. Assuming all sets of locations with at least three faults are malignant, analyzing all pairs of locations inside the CNOT 1-exRec gives for the effective noise strength at level k of the recursive simulation:

$$p_6^{(k)} \leq \sum_{j \leq i=1}^6 \alpha_{ij} p_i^{(k-1)} p_j^{(k-1)} + B_6 (p_{\max}^{(k-1)})^3, \quad (5.12)$$

where $p_{\max}^{(k-1)} = \max_i \{p_i^{(k-1)}\}$ (with our chosen set of CSS operations, $p_{\max}^{(k-1)} = p_6^{(k-1)}$). Third- or higher-order events contribute $B_6 \equiv \binom{297}{3} \leq 4.33 \times 10^6$ and the coefficients α_{ij} when Steane's 1-EC

is used were found to be [3]:

$$\alpha = \begin{pmatrix} 52 & & & & & \\ 174 & 138 & & & & \\ 174 & 0 & 138 & & & \\ 234 & 0 & 378 & 216 & & \\ 234 & 378 & 0 & 0 & 216 & \\ 1154 & 1281 & 1281 & 1566 & 1556 & 3733 \end{pmatrix}. \quad (5.13)$$

Thus, by pessimistically taking all noise strengths equal, we obtain $A_6 = 12913$ malignant pairs, or

$$p_6^{(k)} \leq A'_6 \left(p_6^{(k-1)} \right)^2, \quad (5.14)$$

where $A'_6 = \frac{A_6}{2} \left(1 + \sqrt{1 + \frac{4B_6}{A_6^2}} \right) \leq 13241$. This indicates 7.55×10^{-5} as an improved lower bound on p_{thr} .

This lower bound can be improved further if we use the idea of contracting rectangles discussed in §4.2.4. In particular, we can contract state-preparation 1-exRecs and join them with the succeeding CNOT 1-exRecs and, similarly, we can contract measurement 1-exRecs to join them with the preceding CNOT 1-exRecs. That is, for all levels $k \geq 2$ of the recursive simulation, we need not consider separate preparation or measurement exRecs at level $k-1$. Instead, we can consider level- $(k-1)$ *contracted exRecs* or, simply, *conexRecs* that contain level- $(k-1)$ preparation (respectively, measurement) joined with the succeeding (respectively, preceding) level- $(k-1)$ CNOT operations and the intermediate level- $(k-1)$ EC gadgets *omitted*. We observe that the effective noise strength acting on such conexRecs will be no worse than the effective noise strength acting on CNOT exRecs because (i) error correction following state preparation is superfluous as long as the preparation circuit satisfies properties 2, and (ii) the measurement of logical Pauli operators for CSS codes can be performed transversally on each qubit in the code block and, so, classical error correction on the measurement outcomes can replace quantum error correction preceding the measurements.

Thus, for levels $k \geq 2$ of the recursive simulation, we can effectively set those coefficients in the matrix (5.13) that involve preparation or measurement locations to *zero*. Then, for levels $k \geq 2$, we compute $A_{6,\text{str}} = 4939$, $B_{6,\text{str}} = \binom{153}{3} \leq 5.86 \times 10^5$ and, so, $A'_{6,\text{str}} \leq 5055$. Thus, we obtain the recursion equations

$$\begin{aligned} p_6^{(1)} &\leq A'_6 (p_6)^2, \\ p_6^{(k)} &\leq A'_{6,\text{str}} \left(p_6^{(k-1)} \right)^2, \text{ for } k > 1, \end{aligned} \quad (5.15)$$

which imply the condition $p_6^{(1)} < 1.97 \times 10^{-4}$ and $p_6 < 1.22 \times 10^{-4}$. Hence, 1.22×10^{-4} is our new lower bound on p_{thr} .

If we use Knill's 1-EC gadget instead, the analogous coefficients, α_{ij} , of equation (5.12) are [3]:

$$\alpha = \begin{pmatrix} 60 & & & & & \\ 60 & 30 & & & & \\ 60 & 0 & 30 & & & \\ 180 & 0 & 180 & 270 & & \\ 180 & 180 & 0 & 0 & 270 & \\ 1104 & 552 & 552 & 1656 & 1656 & 4164 \end{pmatrix}. \quad (5.16)$$

Taking again all noise strengths equal, $A_6 = 11184$ and we find $A'_6 \leq 11559$. For levels $k \geq 2$ of the recursive simulation, we can again use conexRecs; we then compute $A_{6,\text{str}} = 5328$. Furthermore, by using our observation in §5.2.1.3, we can take the transversal Bell measurements in the *leading* Knill's 1-EC gadgets inside the CNOT exRec to be ideal. This reduces the effective number of locations in this exRec from 297 to $297 - 2 \times 27 = 243$ at level 1, and from 153 to $153 - 2 \times 9 = 135$ at levels $k \geq 2$. The analogous recursion equations to equation (5.15) give 1.26×10^{-4} as our lower bound on p_{thr} in this case.

5.3.3.2 The 25-Qubit Bacon-Shor Code

The 1-EC Gadgets

Syndrome measurements using the gauge qubits for the 25-qubit Bacon-Shor code, $\mathcal{C}_{BS}^{(5)}$, can be implemented with the circuits in figure 5.10; for this code as well, some additional measurements to those shown must be performed in order to satisfy properties 1. If instead we use Steane's 1-EC gadget, we need to prepare five-qubit cat states as is, e.g., shown in figure 5.11. This time the cat states need verification to prevent a single fault from causing more than one X or Z error that will propagate to the subsequent transversal measurements. Alternatively, verification can be replaced by repeated syndrome extraction.

Accuracy Threshold Lower Bounds

Now $t = 2$ and, to lowest nontrivial order, we need to perform an analysis of malignant triples of locations inside the CNOT 1-exRec. Because of the size of this 1-exRec, the combinatorial analysis is significantly more time demanding than for $\mathcal{C}_{BS}^{(3)}$. The analysis in [3] was performed by using a computer program running on a cluster of 20 Pentium III processors for approximately three months. The CNOT 1-exRec we analysed used Steane's 1-EC gadget, with cat-state verification as in figure 5.11. Because this verification is nondeterministic, there is a statistical fluctuation in the number of cat-state preparation attempts that will need to be made before each cat state is accepted. For *stochastic* local noise, we can account for this nondeterminism in our estimate of the accuracy

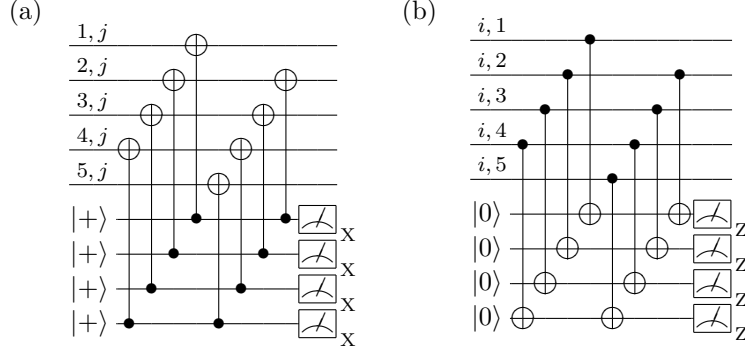


Figure 5.10. A fragment of the 1-EC gadget for $\mathcal{C}_{BS}^{(5)}$ where the syndrome is measured using the gauge qubits. (a) Measuring operators in T along column j . (b) Measuring operators in T along row i . Identical circuits are run in parallel in all rows and columns. Some additional measurements (not shown) are necessary in order to satisfy properties 1; e.g., the syndrome extraction could be repeated.

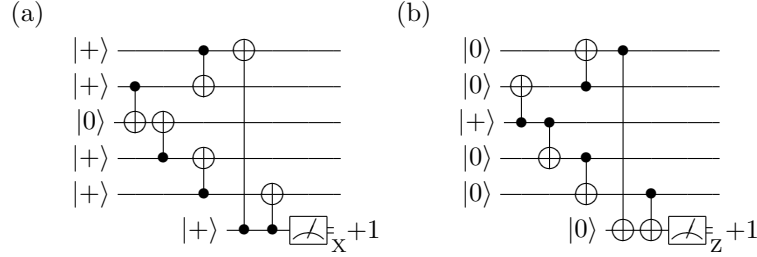


Figure 5.11. (a) A five-qubit cat state in the Hadamard-rotated basis with verification against Z errors. (b) A five-qubit cat state with verification against X errors.

threshold by considering the probability of failure at level k of the recursive simulation *conditioned on* having accepted all cat states inside all error-correction gadgets at *all* levels of the recursion.

The effective noise strength at the level- k CNOT exRec if all of the verification tests contained inside this exRec are successful is

$$p_{6,\text{joint}}^{(k)} \leq \sum_{\ell \leq j \leq i=1}^6 \alpha_{ij\ell} p_i^{(k-1)} p_j^{(k-1)} p_\ell^{(k-1)} + B_6 (p_{\max}^{(k-1)})^4, \quad (5.17)$$

where $B_6 = \binom{1185}{4}$ gives the number of combinations of four locations inside the CNOT 1-exRec which contains 1185 locations in total. We now need to renormalize this noise strength to account for the fact that verification may fail. The probability of accepting all cat states inside Steane's 1-EC is

$$P_{\text{accept}}^{(k)} \geq \left(1 - p_{\max}^{(k-1)}\right)^{C_0}, \quad (5.18)$$

where $C_0 = 190$ is the total number of locations in the encoding and verification circuit for all cat states contained in this 1-EC gadget. Taking the noise strengths for all locations to be equal,

the effective noise strength at the level- k CNOT exRec *conditioned on* the successful outcome in all verification tests contained inside it can be upper bounded by

$$p^{(k)} \leq \left(1 - p^{(k-1)}\right)^{-4C_0} (A_6 + B_6 p^{(k-1)})(p^{(k-1)})^3, \quad (5.19)$$

where from our combinatorial analysis [3], $A_6 = 16625488$ is the number of malignant triples, and we have used the fact that the CNOT 1-exRec contains four 1-EC gadgets. Finally, as for $\mathcal{C}_{BS}^{(3)}$, for all levels $k \geq 2$ of the recursive simulation, we can consider contracted extended rectangles. Then, for levels $k \geq 2$ we only keep the coefficients $\alpha_{ij\ell}$ that do not involve preparation or measurement locations. Since $\alpha_{111} = 287460$, $\alpha_{116} = 1899918$, $\alpha_{166} = 3911460$, and $\alpha_{666} = 2554190$, we have $A_{6,\text{str}} = 8653028$. Also, $B_{6,\text{str}} = \binom{705}{4}$ and $C_{0,\text{str}} = 120$. Thus, by solving for the fixed point of our recursion equations, we obtain 1.94×10^{-4} as our lower bound on p_{thr} .

5.3.3.3 Summary of Results

Table 5.1 summarizes our results [3]. As already discussed, we carried out the combinatorial analysis for the concatenated $\mathcal{C}_{BS}^{(3)}$ using both Steane's and Knill's 1-EC gadget and for the concatenated $\mathcal{C}_{BS}^{(5)}$ using Steane's 1-EC gadget. Our best lower bound on the accuracy threshold for stochastic local noise is 1.94×10^{-4} and it was obtained with the concatenated $\mathcal{C}_{BS}^{(5)}$; this lower bound improves by nearly an order of magnitude the 2.73×10^{-5} rigorous lower bound established in [1] with the concatenated Steane $[[7,1,3]]$ code.

Code	Parameters	1-EC gadg.	exRec locs.	$p_{\text{thr}} (\times 10^{-4})$	$p_{\text{thr}}^{\text{MC}} (\times 10^{-4})$
Steane	$[[7,1,3]]$	Steane	575	0.27	
$\mathcal{C}_{BS}^{(3)}$	$[[9,1,3]]$	Steane	297	1.21	1.21 ± 0.06
		Knill	297	1.26	1.26 ± 0.05
$\mathcal{C}_{BS}^{(5)}$	$[[25,1,5]]$	Steane	1,185	1.94	1.92 ± 0.02
		Knill	1,185		2.07 ± 0.03
Golay	$[[23,1,7]]$	Steane	7,551		≈ 1
$\mathcal{C}_{BS}^{(7)}$	$[[49,1,7]]$	Steane	2,681		1.74 ± 0.01
		Knill	2,681		1.91 ± 0.01

Table 5.1. Rigorous lower bounds on the accuracy threshold, p_{thr} , for stochastic local noise with the concatenated Bacon-Shor code of varying block size and comparison with prior rigorous lower bounds using the concatenated Steane $[[7,1,3]]$ code [1] and Golay $[[23,1,7]]$ code [92]. The fourth column gives the number of locations in the CNOT 1-exRec. The fifth column gives our lower bounds on p_{thr} . The last column gives Monte-Carlo estimates for p_{thr} with 1σ uncertainties. Bold fonts indicate the best results in each column.

Analyzing codes of larger block size than $\mathcal{C}_{BS}^{(5)}$ proved to be computationally difficult in this exact setting. In these cases, we may use a Monte-Carlo method and uniformly sample the set of fault paths with a fixed number of faulty locations inside an extended rectangle. By estimating what

fraction, \hat{f} , of these sets is malignant, we obtain an estimate of the exact combinatorial coefficients with a standard error that can be determined by using $\sigma_{\hat{f}} = \sqrt{\hat{f}(1 - \hat{f})/N}$ where N is the sample size. We have also applied this Monte-Carlo method to the cases where we could extract the exact combinatorics in order to provide evidence that the Monte-Carlo estimates are accurate. Indeed, as can be seen in table 5.1, the exact lower bounds in those cases lie within 1σ of the estimated lower bounds.

5.3.4 Universal Quantum Computation

It remains to discuss the fault-tolerant simulation of non-CSS operations that are required for quantum universality. We will see that as long as the physical noise strength is below the threshold for CSS operations, it will be possible to obtain arbitrarily reliable simulations of these additional operations as well. Therefore, the lower bounds presented in the previous section do indeed apply to universal quantum computation.

Our first step will be to add to the set of CSS operations the phase gate, S , that allows us to generate the Clifford group. Next, to complete quantum universality, we will further add the non-Clifford Toffoli gate. Both the logical phase and Toffoli gates will be simulated by using CSS operations and by consuming special “quantum software” ancillary states; §5.3.4.1 discusses the simulation of the logical phase gate and §5.3.4.2 the simulation of the logical Toffoli gate. In both these sections, we will assume we are able to prepare these ancillary states with a fault probability that is upper bounded by some constant, p_{anc} , and we will discuss how arbitrarily accurate copies can be distilled from them if p_{anc} is below some distillation threshold value. We will postpone until the final section, §5.3.4.3, the discussion about how ancillary states with accuracy below this distillation threshold can be prepared for the 9- and 25-qubit Bacon-Shor codes.

5.3.4.1 The Logical Phase Gate

We first note that since the logical CNOT and Hadamard gates are transversal, the logical CPHASE is also transversal; it can be implemented by doing a logical Hadamard on the target, followed by a logical CNOT, followed by another logical Hadamard on the target.

A destructive measurement of the logical X (respectively, Z) operator can be performed by transversally measuring the operator X (respectively, Z) on each qubit in the code block. (With the logical CNOT and CPHASE transversal, we can also measure non-destructively the logical X or the logical Z operator by using as control an ancilla prepared in the logical $|+\rangle$ state.) If we were also able to easily measure the logical Y , $Y_L = iX_L Z_L$, we would be able to implement the logical phase gate by using, e.g., the circuit in figure 5.12.

Although $Y_L = \otimes_{i=1}^{n^2} Y^{(i)}$ when n is odd, measuring transversally the operators $Y^{(i)}$ does not

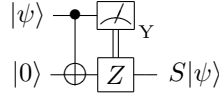


Figure 5.12. Circuit simulation of the phase gate, S , by using a measurement of Y . Contrary to the standard notation, here the correction operator Z is applied when the measurement outcome is $+1$ and no correction is applied when the outcome is -1 .

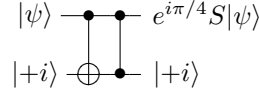


Figure 5.13. Circuit simulation of the phase gate S using the ancillary state $|+i\rangle$.

give a measurement 1-Ga which satisfies property 2(b). This is because the Bacon-Shor code has no stabilizer operators that can be written as tensor products of Y operators *alone*. Therefore, the problem is that we cannot perform error correction on the transversal measurement outcomes and, so, the eigenvalue we would deduce for the logical Y could be erroneous even if a single one of the transversal measurements failed.

We could instead measure the logical Y operator nondestructively using cat states similar to our discussion in §5.2.1.1. Implementing this measurement would require controlled- Y gates, which at the next level of the recursive simulation would have to be implemented in an encoded form. But then, the problem is that the controlled- Y gate is complex and all the transversal operations we have discussed so far (CNOT, CPHASE and Hadamard) are real. Therefore, we do not have a direct transversal method for implementing the logical controlled- Y gate.

Fortunately, there is a method for simulating the logical S gate by using only logical CNOT and CPHASE gates provided we can prepare a certain logical ancillary state: Consider the states $|\pm i\rangle \propto |0\rangle \pm i|1\rangle$ which are the ± 1 eigenstates of Y . Given the state $|+i\rangle$, we can simulate S (up to an irrelevant phase) with the circuit in figure 5.13. Hence, the problem of constructing a 1-Ga for the logical phase gate reduces to the problem of preparing the *logical* ancillary state $|+i\rangle$.

Simulating S and S^* in Superposition

Based on the circuit in figure 5.13, we will now first describe a simple procedure by which the simulation of the logical S gate is still possible even if the logical ancillary state $|+i\rangle$ is replaced by another logical ancillary state which is easier to prepare [93]. This procedure makes use the following two observations: (i) the S gate is the only complex gate in our universal gate set, and (ii) if the state $| -i\rangle$ is used instead in the the circuit in figure 5.13, then the operation $S^\dagger = S^*$ will be simulated.

The idea is to use some standard state, say the state $|0\rangle$, instead of $|+i\rangle$ in the circuit in figure

5.13. We can expand $|0\rangle \propto |+i\rangle + |-i\rangle$ and, for each of the two terms, we can consider the two paths of the subsequent computation which are executed in superposition. In one path S is simulated and in the other S^* . Thus, if every time we want to simulate S in our circuit we use the *same* ancillary $|0\rangle$ state and because S is the only complex gate in our gate set, the final state of the computation will be a linear superposition of one term where the desired computation unitary, U , has been implemented and a second term where U^* has been implemented instead. In other words, if the initial computation state is $|\psi_{\text{initial}}\rangle$, then the final computation state will be

$$|\psi_{\text{final}}\rangle = \frac{|+i\rangle \otimes U|\psi_{\text{initial}}\rangle + |-i\rangle \otimes U^*|\psi_{\text{initial}}\rangle}{\sqrt{2}} . \quad (5.20)$$

In the end of the computation some operator A will be measured which we can take to be real (and so, due to hermiticity, $A^T = A$). We now want to see that the expectation value for A will be the same as if the desired U had been simulated all along. Indeed, we compute

$$\langle A \rangle = \langle \psi_{\text{initial}} | \frac{U^\dagger A U + (U^*)^\dagger A U^*}{2} | \psi_{\text{initial}} \rangle = \langle \psi_{\text{initial}} | U^\dagger A U | \psi_{\text{initial}} \rangle , \quad (5.21)$$

since, $\forall |\psi\rangle$, $\langle \psi | (U^*)^\dagger A U^* | \psi \rangle = \langle \psi | U^\dagger A^T U | \psi \rangle = \langle \psi | U^\dagger A U | \psi \rangle$.

We can use this procedure at the logical level as well. The only penalty we pay for simulating the logical S gate in this way is that we need to swap around the ancillary logical $|0\rangle$ block we use for the simulation if we are constrained to use only local interactions; but this will only give us a linear penalty in the size of the computation. Of course, we should emphasize that this trick works because the S gate is not used in implementing error correction. As a consequence, we only need to implement logical S gates at the *highest* level of the recursive fault-tolerant simulation and, moreover, it is not necessary that we execute different logical S gates in parallel anywhere in our computation. Therefore, provided the physical noise strength is below the accuracy threshold for CSS operations, we may obtain any desired accuracy in the simulation of the logical S gates at the highest level of our recursive simulation by choosing this highest level appropriately.

Noisy $|+i\rangle$ Distillation

In cases when S is not the only complex gate in our gate set, there is also a straightforward procedure for fault-tolerantly preparing the required ancillary “quantum software.” We can begin by preparing many noisy logical $|+i\rangle$ states and use them to progressively distill less and less noisy copies. A possible distillation circuit is shown in figure 5.14.

The measurement outcome in this circuit is ideally $+1$, and an error on *one* of the two $|+i\rangle$ states can be detected. We note that because Y stabilizes the state $|+i\rangle$, we need only worry about Z errors—we can write $X = iZY$ which is, up to the irrelevant phase i , equivalent to Z when it acts on $|+i\rangle$. Postselecting on the $+1$ measurement outcome, the fidelity of the output $|+i\rangle$ state is

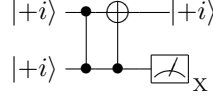


Figure 5.14. Circuit for the nondeterministic distillation of the state $|+i\rangle$. Postselecting on the $+1$ measurement outcome, the output $|+i\rangle$ state has quadratically improved fidelity.

increased quadratically relative to the fidelity of the input state since it takes errors in *both* input $|+i\rangle$ states for an error in the output state to go undetected.

To be concrete, assume noise is local and stochastic and let p_{anc} denote the probability that some fault has occurred during the preparation of one of the initial copies of the distillation protocol. Then by “twirling,” i.e., by applying at random with probability $1/2$ a Y operator (and by doing nothing with probability $1/2$), we may describe the initial copies as having a Z error with probability at most p_{anc} . If the CSS operations of the protocol are executed ideally, then the probability of a Z error on the output copy conditioned on getting the $+1$ measurement outcome and accepting the state is

$$p_{\text{anc}}^{(1)} \leq \frac{p_{\text{anc}}^2}{p_{\text{anc}}^2 + (1 - p_{\text{anc}})^2} \leq 2p_{\text{anc}}^2, \quad (5.22)$$

since both input copies must have a Z error in order for an erroneous output state to be accepted.⁴

If we repeat the distillation protocol $l \geq 2$ times and since the input copies for each round are prepared independently at the previous round, we find $p_{\text{anc}}^{(l)} \leq \frac{1}{2} (2p_{\text{anc}})^{2^l}$ which indicates a distillation threshold, $p_{\text{thr}}^{\text{dist}, |+i\rangle} = 1/2$.

5.3.4.2 The Logical Toffoli Gate

The simulation of the logical Toffoli gate uses the logical ancillary state

$$|\text{Toffoli}\rangle \equiv \Lambda^2(X) (|+\rangle \otimes |+\rangle \otimes |0\rangle)$$

as an input to the circuit in figure 5.15 [51, 67]. Depending on the ± 1 measurement outcomes, the post-measurement Clifford-group corrections are $S_1^{k_1} S_2^{k_2} S_3^{k_3}$ where $k_i = 0$ if $j_i = +1$ and $k_i = 1$ if $j_i = -1$; here,

$$S_1 = X^{(1)} \otimes \text{CNOT}_{2 \rightarrow 3}, S_2 = X^{(2)} \otimes \text{CNOT}_{1 \rightarrow 3}, \text{ and } S_3 = \text{CPHASE}_{1,2} \otimes Z^{(3)},$$

and let us note that these operators are mutually commuting and $|\text{Toffoli}\rangle$ is their simultaneous $+1$ eigenvector. Therefore, the problem of constructing a 1-Ga for the logical Toffoli gate reduces to preparing a logical ancillary state, the logical $|\text{Toffoli}\rangle$.

⁴Here, we used the inequality, $\forall \eta_1, \eta_2 \leq 1/2, \eta_1 \eta_2 + (1 - \eta_1)(1 - \eta_2) \geq 1/2$, which also implies that the bound (5.22) is valid even if the error probabilities on the two input copies are not exactly equal.

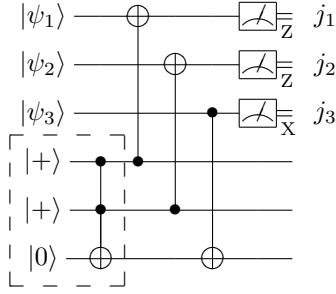


Figure 5.15. Circuit simulation of the Toffoli gate using the ancilla $|\text{Toffoli}\rangle$ shown in a box. Depending on the measurement outcomes, Clifford-group corrections (not shown) may need to be applied to the output qubits. To see that the circuit teleports the Toffoli gate, note that the Toffoli gate inside the box can be commuted to the right of the circuit thereby leaving behind three teleportation circuits for each of the three input qubits.

Recursive $|\text{Toffoli}\rangle$ Preparation

Let us consider starting with the state $|+\rangle^{\otimes 3}$ which is clearly a $+1$ eigenstate of both S_1 and S_2 ; then, to obtain the state $|\text{Toffoli}\rangle$, it remains to perform a measurement of the operator S_3 , e.g., with the circuit in figure 5.16 [67].

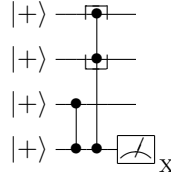


Figure 5.16. A circuit that projects the input state $|+\rangle^{\otimes 3}$ onto the $+1$ eigenstate of the operator S_3 depending on the measurement outcome. For outcome $+1$ the output state is $|\text{Toffoli}\rangle$.

When this circuit is realized in an encoded form, the first three qubits become three code blocks each encoded in the logical $|+\rangle$ state and the fourth qubit is replaced by n_c qubits in a cat state, where n_c is the number of qubits in the code block. Since the measurement outcome can be erroneous due to even a single fault acting on the cat-state qubits, the measurement of the *logical* S_3 must be repeated several times to ensure properties 2 are satisfied. If the code has distance $d = 2t + 1$, d repetitions are sufficient since then the majority vote is guaranteed to give the correct outcome if no more than t of the individual measurement outcomes are erroneous, thereby reducing the probability of failure to order $t + 1$. Furthermore, in order to correct faults in the noisy Toffoli gates, error-correction steps must be inserted on each one of the three code blocks between successive measurements.

Let us also comment on the second gate in figure 5.16. Since the logical CPHASE is transversal, applying the logical CPHASE gate on the first two code blocks with a *control* by the cat-state qubits

can be implemented as follows: We may first rotate the qubits in, say, the first code block by 90 degrees; next, we apply controlled-CPHASE gates bitwise with control the cat state and target the first two code blocks; finally, we rotate back by 90 degrees the qubits in the first code block. By conjugating with bitwise Hadamard gates the cat-state qubits, the controlled-CPHASE gates become Toffoli gates so that the entire preparation circuit for the $|\text{Toffoli}\rangle$ state uses only CSS operations and Toffoli gates. This is important because it allows us to apply this simulation procedure recursively. That is, we may start with “level-0” $|\text{Toffoli}\rangle$ states that enable the simulation of noisy “level-0” Toffoli gates and, next, we may use these gates to prepare “level-1” $|\text{Toffoli}\rangle$ states that will allow us to simulate less noisy “level-1” Toffoli gates, etc., until we eventually quickly reach our desired accuracy.

To be more concrete, let p_{anc} be the probability that a fault has occurred during the preparation of a “level-0” $|\text{Toffoli}\rangle$ state, maximized over all our copies of such states. If we assume all CSS operations in the circuit in figure 5.16 are executed ideally, the noise strength on the output “level-1” $|\text{Toffoli}\rangle$ state is

$$p_{\text{anc}}^{(1)} \leq d \binom{n_c}{t+1} p_{\text{anc}}^{t+1} + \binom{d}{t+1} (n_c p_{\text{anc}})^{t+1}. \quad (5.23)$$

Here, the first term is due to the fact that at least $t+1$ faults must have occurred in at least one of the d sequential executions of the circuit in figure 5.16 in order for a logical error to be created—any number of faults less than $t+1$ will be corrected by the intermediate error-correction steps that are executed ideally since they only contain CSS operations. The second term is due to the possibility of having at least $t+1$ of the d measurements of the logical S_3 give erroneous outcomes—for each individual measurement of S_3 , an error can be produced due to a fault in any one of the n_c Toffoli gates inside the measurement circuit.

By repeating this simulation procedure recursively, we may achieve any desired accuracy for the simulation of the logical Toffoli gate as long as the noise strength acting on the initial “level-0” $|\text{Toffoli}\rangle$ state is below a threshold value; equation (5.23) implies that this threshold value is

$$p_{\text{thr,determ}}^{\text{dist},|\text{Toffoli}\rangle} = \left(d \binom{n_c}{t+1} + \binom{d}{t+1} n_c^{t+1} \right)^{-1/t}. \quad (5.24)$$

For the $[[n_c = 9, k = 1, d = 3]]$ Bacon-Shor code we find a threshold of at least .28%; for the $[[n_c = 25, k = 1, d = 5]]$ Bacon-Shor code, .24%.

Noisy $|\text{Toffoli}\rangle$ Distillation

Alternatively, we can consider preparing several copies of the logical $|\text{Toffoli}\rangle$ state and using them to distill less noisy output copies [94, 24]. Let us first construct a circuit that takes as an input some number of copies of the $|\text{Toffoli}\rangle$ state and outputs one $|\text{Toffoli}\rangle$ state of improved fidelity.

Consider the circuit in figure 5.17(a). We start with one copy of the state $|\text{Toffoli}\rangle$ and we measure

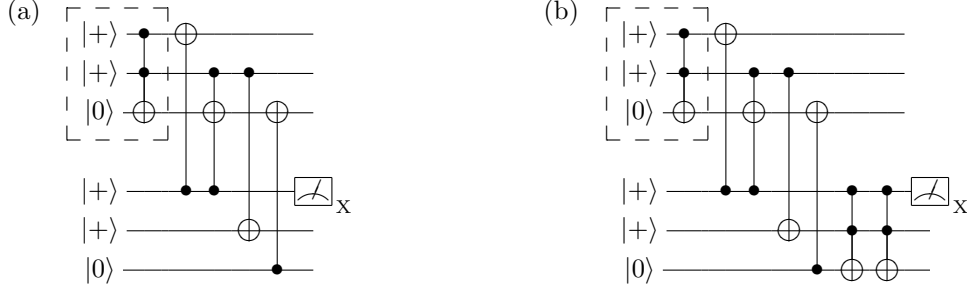


Figure 5.17. (a) A circuit that implements a measurement of S_1 on the input $|\text{Toffoli}\rangle$ state shown in a box. The measurement outcome is ideally +1 since $|\text{Toffoli}\rangle$ is an eigenstate of S_1 . Note that the gates acting on the last two qubits operate trivially. (b) The same circuit after a pair of cancelling Toffoli gates has been inserted.



Figure 5.18. Commutation relations of CNOT and Toffoli.

the operator S_1 using one ancillary qubit initialized in the state $|+\rangle$. Since the state $|\text{Toffoli}\rangle$ is an eigenstate of this operator, the effect of the measurement is trivial: the first three qubits remain in the state $|\text{Toffoli}\rangle$ and the measurement on the ancilla always ideally gives outcome +1. The circuit in figure 5.17 includes two extra ancillary qubits. These qubits are useful because we now want to insert two cancelling Toffoli gates to obtain the equivalent circuit in figure 5.17(b). Our motivation for inserting the two Toffoli gates is to create two copies of the $|\text{Toffoli}\rangle$ state in the input. Thus, we next commute the leftmost of the two Toffoli gates to the left by using repeatedly the circuit identities in figure 5.18. After the dust settles, the result is the circuit shown in figure 5.19(a).

This circuit includes, except for the two input $|\text{Toffoli}\rangle$ states, additional Toffoli gates acting between these two states that we now want to eliminate—recall that our goal is to design a distillation circuit for the $|\text{Toffoli}\rangle$ state that uses exclusively CSS operations. We observe that if we measure the last two qubits as shown in figure 5.19(b), we can apply corrections depending on the measurement outcomes that will restore the state of the first three qubits to be the $|\text{Toffoli}\rangle$ state. Furthermore, if we subsequently measure the fourth qubit, the eigenvalue will correspond to the *product* of the eigenvalues of S_1 acting on the two input $|\text{Toffoli}\rangle$ states⁵. Ideally this product is +1 because both input $|\text{Toffoli}\rangle$ states are eigenstates of S_1 . On the other hand, if the measured product is -1 , we know that one of the input $|\text{Toffoli}\rangle$ states had an error. Hence, if we postselect on the +1

⁵To see this, note that in figure 5.17 the measurement gives the eigenvalue of $S_1 \otimes X$ where S_1 acts on the first three qubits and X on the fourth qubit. By commuting X to the right of the Toffoli gate inside the preparation circuit of the second $|\text{Toffoli}\rangle$ copy, we see that we equivalently implement a measurement of the operator $S_1 \otimes S_1$ on the two input $|\text{Toffoli}\rangle$ states.

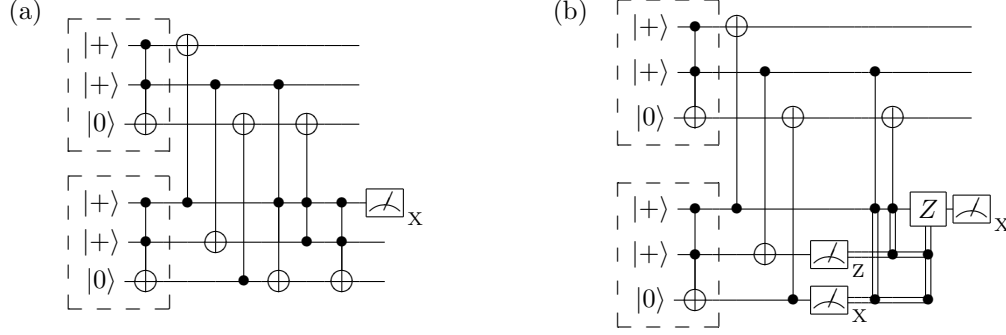


Figure 5.19. (a) An equivalent circuit to those in figure 5.17. We identify the input to be two copies of the state $|\text{Toffoli}\rangle$ shown in boxes. (b) After measuring the last two qubits and applying the appropriate corrections, the first three qubits are left in the state $|\text{Toffoli}\rangle$. The fourth qubit is then measured and the outcome gives the product of the eigenvalues of S_1 acting in the two input $|\text{Toffoli}\rangle$ states.

measurement outcome, we expect the fidelity of the output $|\text{Toffoli}\rangle$ state to improve quadratically.

In more detail, the 2^3 eigenstates of the three commuting operators S_1 , S_2 and S_3 form an orthonormal basis in the three-qubit Hilbert space. We can label the basis states by $|\text{Toffoli}\rangle_{a,b,c}$ depending on their eigenvalues, $a, b, c = \pm 1$, with respect to the operators S_1 , S_2 and S_3 . Then, $|\text{Toffoli}\rangle \equiv |\text{Toffoli}\rangle_{+1,+1,+1}$ and we note that

$$|\text{Toffoli}\rangle_{a,b,c} = Z^{\bar{a}} \otimes Z^{\bar{b}} \otimes X^{\bar{c}} |\text{Toffoli}\rangle, \quad (5.25)$$

where $\bar{a} = 0$ if $a = +1$ and $\bar{a} = 1$ if $a = -1$, and similarly for \bar{b} and \bar{c} . Let us assume that p_{anc} is an upper bound on the probability of a fault during the preparation of a $|\text{Toffoli}\rangle$ state that is input to the distillation protocol. Then, by flipping three coins and applying with probability $1/2$ each of S_1 , S_2 , and S_3 (and doing nothing otherwise in each of the three cases), we may assume that the input copies to the protocol are mixtures of the basis states $\{|\text{Toffoli}\rangle_{a,b,c}\}$ where the probability of the state $|\text{Toffoli}\rangle$ in the mixture is at least $1 - p_{\text{anc}}$; we may also say that the “errors” $Z \otimes I \otimes I$, $I \otimes Z \otimes I$, and $I \otimes I \otimes X$ appear with probabilities p_a , p_b , and p_c such that $p_a, p_b, p_c \leq p_{\text{anc}}$.

Let us now consider the circuit in figure 5.19(b) and let the two input states to the distillation be the states $|\text{Toffoli}\rangle_{a,b,c}$ and $|\text{Toffoli}\rangle_{a',b',c'}$. In such a case the first three qubits at the output are in the state $|\text{Toffoli}\rangle_{a,b,b',c,c'}$ and, moreover, the measurement outcome on the fourth qubit corresponds to $a \cdot a'$, i.e., it equals the product of the eigenvalues of S_1 of the two input copies. Therefore, postselecting on the $+1$ measurement outcome, we achieve a quadratic improvement of the error probability p_a in the surviving state $|\text{Toffoli}\rangle_{a,b,b',c,c'}$,

$$p_a \rightarrow p'_a \leq \frac{p_a^2}{p_a^2 + (1 - p_a)^2} \leq 2p_a^2. \quad (5.26)$$

On the other hand, conditioned on accepting the output state, the output probabilities p_b and p_c get worse since if any one of the input copies was a -1 eigenstate of S_2 or S_3 , the surviving copy will also be such a -1 eigenstate. In particular,

$$p_b \rightarrow p'_b \leq \frac{2p_b(1-p_b)}{p_a^2 + (1-p_a)^2} \leq 4p_b, \quad (5.27)$$

$$p_c \rightarrow p'_c \leq \frac{2p_c(1-p_c)}{p_a^2 + (1-p_a)^2} \leq 4p_c. \quad (5.28)$$

Because the error probabilities with respect to the parameters b and c only get worse by a linear factor, if we repeat the distillation protocol using a different circuit that purifies with respect to b and also with a third circuit that purifies with respect to c , we will eventually achieve a quadratic improvement in all three probabilities p_a , p_b , and p_c [94, 24]⁶. Overall, after these three distillation subprotocols, the error probabilities evolve as $(p_a, p_b, p_c) \rightarrow (32p_a^2, 128p_b^2, 512p_c^2)$, thus indicating a threshold value of at least $1/512 \geq .19\%$.

In fact, we can improve this value if we note that provided the error parameters never exceed 12%, we may use the bound $p_i^2 + (1-p_i)^2 \geq .75$ to replace equations (5.26) to (5.28) by $(p_a, p_b, p_c) \rightarrow (\beta p_a^2, \gamma p_b^2, \gamma p_c^2)$ where $\beta = \frac{4}{3}$ and $\gamma = \frac{8}{3}$. After the three distillation subprotocols, the error probabilities evolve as

$$(p_a, p_b, p_c) \rightarrow (\beta\gamma^2 p_a^2, \beta\gamma^3 p_b^2, \beta\gamma^4 p_c^2), \quad (5.29)$$

indicating a distillation threshold,⁷ $p_{\text{thr}}^{\text{dist}, |\text{Toffoli}\rangle} \geq (\beta\gamma^4)^{-1} \geq 1.45\%$.

5.3.4.3 Bounds on the Accuracy of Quantum Software

Let us finally discuss how we can prepare copies of the $|+i\rangle$ and $|\text{Toffoli}\rangle$ states that are used as inputs to the distillation protocols described in the previous two sections. There are two issues to be addressed: The first one is related to the fact that our distillation procedures assumed that CSS operations were executed without faults. Although this assumption is never exactly satisfied, our idea was that we can *simulate* CSS operations to any desired accuracy provided the physical noise strength is below the accuracy threshold for CSS operations. Therefore our assumption about noiseless CSS operations is justified if we simulate the distillation protocols at a sufficiently high level, say k , of a recursive fault-tolerant simulation; this then implies that the input ancillary states to the distillation protocols must also be encoded in a level- k code block. The second issue relates to the accuracy of these level- k ancillary states: Not only must we show how such states can be

⁶Such circuits can be easily constructed using a similar trick as the one we used here to derive the circuit in figure 5.19(b) from that in figure 5.17.

⁷We may check for consistency that p_c after the second subprotocol remains below our 12% upper bound; indeed, $\gamma^2 \times 1.45\% < 12\%$.

prepared, but we must also be able to execute this preparation in such a way that the probability of a fault during preparation is below the distillation threshold.

Perhaps the simplest method for preparing a level- k code block in a desired single-qubit state, $|\psi\rangle$, is to use quantum teleportation: We prepare a single qubit in the state $|\psi\rangle$ and some ancillary qubits in the state $|0\rangle|0\rangle_L + |1\rangle|1\rangle_L$ where, e.g., $|0\rangle|0\rangle_L$ denotes the $|0\rangle$ state on the first ancillary qubit and the level- k logical $|0\rangle$ state on the remaining ancillary qubits. By performing a measurement along the Bell basis between the qubit in the state $|\psi\rangle$ and the first ancillary qubit, we teleport the level- k $|\psi\rangle$ state to the remaining ancillary qubits. In chapter B in the appendix we give more details on this procedure and we also derive the following upper bound on the accuracy of the level- k code block,

$$p_{\text{anc}} \leq 3p^{(k)} + D \sum_{j=0}^{k-1} p^{(j)} + 4p, \quad (5.30)$$

where $p^{(k)}$ is the noise strength for CSS operations at level k of the recursion ($p^{(0)} = p$ is the physical noise strength) and D is the number of locations in a decoding circuit that performs a canonical mapping of the state of a level-1 code block to the state of one qubit⁸. For the three-qubit Toffoli state, we can similarly teleport the state of each qubit to a different code block. Then, we can upper bound the probability of a fault that is supported on a specific one of the three level- k code blocks as in equation (5.30) with $4p$ replaced by $7p$.

To show that our accuracy thresholds for CSS operations in §5.3.3.3 do indeed apply to universal quantum computation, it remains to evaluate the upper bound in equation (5.30) for our “quantum software” ancillary states and show that it is below the distillation thresholds.

The 9-Qubit Bacon-Shor Code

We first note that the first term in equation (5.30) can be ignored since it can become arbitrarily small by choosing k sufficiently large provided p is below the accuracy threshold for CSS operations. To upper bound the second term, we recall the recursion equations (5.15) that relate the accuracy of the level- k simulation to the accuracy at level $k-1$. We have

$$D \sum_{j=0}^{k-1} p^{(j)} \leq D \left(p + A'_6 p^2 + \sum_{j=2}^{\infty} \frac{1}{A'_{6,\text{str}}} (A'_{6,\text{str}} (A'_6 p^2))^{2^{j-1}} \right). \quad (5.31)$$

To determine the value of D we need an explicit construction of a decoding circuit such as the one shown in figure 5.20; for this decoder, $D = 16$.

We can now obtain a numerical upper bound on p_{anc} by explicitly calculating the first, say, 100 terms of the sum in equation (5.31) and upper bounding the contribution of the remaining terms

⁸Here we assume that the same code is used at all levels of the recursive simulation; otherwise, D would need to be maximized over all codes used at different levels.

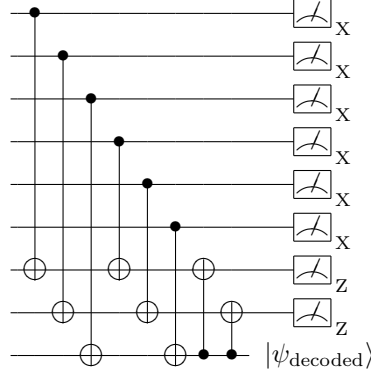


Figure 5.20. A decoding circuit for the 9-qubit Bacon-Shor code. The qubits in the code block are arranged starting with the qubit on the top left corner in the lattice in figure 5.6 and moving from left to right across successive rows.

using

$$\sum_{j=102}^{\infty} p^{(j)} \leq \frac{1}{A'_{6,\text{str}}} \sum_{j=0}^{\infty} \left(A'_{6,\text{str}} p^{(102)} \right)^{2^j} \leq \frac{1}{A'_{6,\text{str}}} \frac{1}{1 - \left(A'_{6,\text{str}} p^{(102)} \right)^2}, \quad (5.32)$$

where we used the bound $\sum_{j=0}^{\infty} x^{2^j} \leq \sum_{j=0}^{\infty} (x^2)^j \leq (1 - x^2)^{-1}$. Using Knill's 1-EC gadget and by substituting the values for A'_6 and $A'_{6,\text{str}}$ and setting $p = p_{\text{thr}} = 1.26 \times 10^{-4}$, we find for single-qubit states, $p_{\text{anc}} < 2.15\%$, which is safely below the 50% $|+i\rangle$ distillation threshold. Similarly, for three-qubit ancillary states, we find that the probability of a fault that is supported on a specific one of the three code blocks is below 2.18%. Since 2.18% is above our 1.45% $|T\text{offoli}\rangle$ distillation threshold, we seem to need to revise our bound on p_{thr} in table 5.1. In fact, decreasing our bound by 10^{-5} to $p_{\text{thr}} = 1.16 \times 10^{-4}$ makes p_{anc} drop to 1.25% for single-qubit states and to 1.29% for three-qubit states which is now below the 1.45% $|T\text{offoli}\rangle$ distillation threshold.

However, it is not necessary that we revise our bound on p_{thr} . The reason why we chose to describe the distillation of $|T\text{offoli}\rangle$ states in §5.3.4.2 is because the distillation procedure was conceptually simple and the distillation threshold condition was easy to derive; however, much higher distillation thresholds are possible if we use more complex protocols for distilling certain “magic” single-qubit states as explained in [97]. In particular, the distillation threshold for the $|H\rangle$ magic state is at least 14% [97] which is significantly above 2.15% showing that if p is below 1.26×10^{-4} universal quantum computation is still possible.

The 25-Qubit Bacon-Shor Code

We may repeat the same calculation by using the recursion equations (5.19) and also⁹ $D = 48$.

⁹A decoding circuit can easily be constructed similar to that in figure 5.20: We connect with CNOT gates all qubits in the first four rows to the qubits in the fifth row ($4 \times 5 = 20$ gates) and, then, we connect with CNOT gates the last qubit in the fifth row to every other qubit in the same row (another 4 gates). Finally, we add 24 single-qubit measurements.

Instead of equation (5.32) we now have

$$\sum_{j=102}^{\infty} p^{(j)} \leq \frac{1}{\Gamma} \sum_{j=0}^{\infty} \left(\Gamma p^{(102)} \right)^{3^j} \leq \frac{1}{\Gamma} \frac{1}{1 - \left(\Gamma p^{(102)} \right)^3}, \quad (5.33)$$

where

$$\Gamma \equiv \frac{A_{6,\text{str}} + B_{6,\text{str}} p^{(102)}}{(1 - p^{(102)})^{4C_{0,\text{str}}}}. \quad (5.34)$$

For Steane's 1-EC gadget and by substituting the values for A_6 , $A_{6,\text{str}}$, B_6 , $B_{6,\text{str}}$, C_0 and $C_{0,\text{str}}$ and setting $p = p_{\text{thr}} = 1.94 \times 10^{-4}$, we find for single-qubit states, $p_{\text{anc}} < 8.49\%$, which is below the 50% $|+i\rangle$ distillation threshold. For three-qubit states, the probability of a fault that is supported on a specific one of the three blocks is below 8.55%. Again, 8.55% is above our 1.45% $|T\text{offoli}\rangle$ distillation threshold but safely below the 14% $|H\rangle$ distillation threshold. Hence, universal quantum computation is possible if p is below 1.94×10^{-4} .

Finally, we should note that the sum $D \sum_{j=0}^{k-1} p^{(j)}$ in equation (5.30) diverges as $k \rightarrow \infty$ when $p = p_{\text{thr}}$ since p_{thr} is defined as the fixed point of our recursion equations (i.e., by definition, if we start with $p = p_{\text{thr}}$ then $\forall j, p^{(j)} = p$). However, the lower bounds we give in table 5.1 are slightly below this fixed point so that in fact the sum is convergent when we set $p = p_{\text{thr}}$.

5.4 History and Acknowledgements

I first heard Bacon talk about the Bacon-Shor code at the QIP 2005 workshop, but it did not occur to me that this code could be useful for fault-tolerant quantum computation until after I read [91] and I understood the relation to Shor's code (which, perhaps surprisingly, is not mentioned in [91]). The analysis of malignant sets of locations for $\mathcal{C}_{BS}^{(3)}$ and $\mathcal{C}_{BS}^{(5)}$ was done by Cross and the results appear in [3].

The idea that the accuracy threshold for stabilizer operations (and CSS operations, in particular) sets the overall threshold for universal quantum computation has appeared in the work of Knill [58, 59] and has also been discussed by Bravyi and Kitaev [97] and by Reichardt [57, 98].

Chapter 6

Epilogue

The most valuable intuitions are the last to be attained; the most valuable of all are those which determine methods.

—Nietzsche, *Antichrist*.

Level reduction is such a simple idea that it now appears to us to be completely natural. But as all valuable intuitions, the insights leading to a syntactic analysis of fault-tolerant simulations and the concept of level reduction have become possible based on a large body of work in the theory of fault-tolerant quantum computation during the last ten years.

The value of level reduction is that it provides a simple methodology for analyzing recursive fault-tolerant simulations. In this thesis, this methodology was used to derive rigorous lower bounds on the accuracy threshold and to extend the proof of the quantum threshold theorem to new directions such as computation in the presence of coherent and leakage noise and measurement-based models of quantum computation. Level reduction is also used in a proof of the threshold theorem for postselected quantum computation [95].

Although the family of noise models I have called *local* includes models which are significantly more realistic than any of those considered in previous proofs, there is still a long road toward understanding and limiting the effect of physical noise on future quantum computers. In particular, it would be important to obtain new formulations of the quantum threshold theorem in terms of quantities such as spectral densities and correlation functions of the environment which are more easily accessible experimentally.

Let us hope that level reduction will become a new helpful tool among the many that will be needed to realize a formidable quantum computer.

Appendix A

Deriving the Bacon-Shor code

A.1 Starting from Shor's Code

One natural way to derive the distance- n Bacon-Shor code, $\mathcal{C}_{BS}^{(n)}$, is by starting with Shor's distance- n code, $\mathcal{C}_{Shor}^{(n)}$, as was done for the special case $n = 3$ in [96].

If we place the n^2 qubits in the $\mathcal{C}_{Shor}^{(n)}$ block on the *vertices* of the $n \times n$ square lattice in figure 5.6, then the code's stabilizer group is

$$\mathcal{S}(\mathcal{C}_{Shor}^{(n)}) = \langle X_{j,*} X_{j+1,*} ; Z_{i,j} Z_{i,j+1} \mid i \in \mathbb{Z}_n ; j \in \mathbb{Z}_{n-1} \rangle . \quad (\text{A.1})$$

Also, recall that the logical Pauli X and Z operators are $X_{1,*}$ and $Z_{*,1}$, respectively.

By its construction, Shor's code treats X and Z errors asymmetrically: In each of the n rows, up to $\lfloor n/2 \rfloor$ X errors can be corrected because of the underlying classical repetition code. The code can also correct up to $\lfloor n/2 \rfloor$ Z errors on different rows—pairs of Z operators in the same row act trivially; the code is degenerate. However, we observe that pairs of X operators in the same *column* commute with the logical Z operator. It would therefore be sufficient for successful error correction if we could restore zero parity for X errors in each given column instead of correcting X errors in each row separately.

Since only the parity of X errors in each column is relevant, we can replace the n stabilizer generators $Z_{i,j} Z_{i,j+1}$ for $i \in \mathbb{Z}_n$ and *fixed* j by their tensor-product, $Z_{*,j} Z_{*,j+1}$. Repeating for all $j \in \mathbb{Z}_{n-1}$, we reduce the stabilizer group of $\mathcal{C}_{Shor}^{(n)}$ in equation (A.1) to the stabilizer group of $\mathcal{C}_{BS}^{(n)}$ in equation (5.4).

A.2 Starting from Bravyi and Kitaev's Surface Code

We will also give an alternative derivation of $\mathcal{C}_{BS}^{(n)}$ starting from Bravyi and Kitaev's distance- n surface code, $\mathcal{C}_{BK}^{(n)}$ [99]. If we place the $n^2 + (n-1)^2$ qubits in the $\mathcal{C}_{BK}^{(n)}$ block on the *edges* of a square

lattice as shown in figure A.1, the code's stabilizer group is generated by X operators acting on all qubits neighboring a vertex (*site operators*) and Z operators acting on all qubits neighboring a vertex of the *dual* lattice (*plaquette operators*). The logical Pauli X and Z operators are $X_L = X_{1,*}$ and $Z_L = Z_{*,1}$, respectively. It can easily be verified that if all qubits with half-integer coordinates are measured in the computation basis, the remaining qubits are left in a $+1$ eigenstate of the stabilizer generators in equation (5.4). Moreover, the logical X and Z operators have no support on the measured qubits and match the logical Pauli operators for $\mathcal{C}_{BS}^{(n)}$. This implies that the same logical state that was encoded in the $\mathcal{C}_{BK}^{(n)}$ block is, after the measurements, encoded in the protected qubit of $\mathcal{C}_{BS}^{(n)}$.

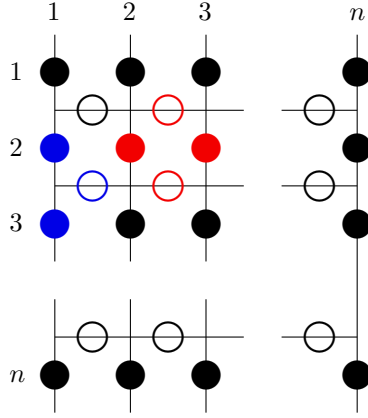


Figure A.1. Qubits in the $\mathcal{C}_{BK}^{(n)}$ block sit on the *edges* of a square lattice with different boundary conditions at top-bottom (*rough* edges) and left-right (*smooth* edges). Two elements of the code stabilizer are shown: X is applied on qubits shown in blue (*site operator*); Z is applied on qubits shown in red (*plaquette operator*). If all qubits shown as empty circles are measured in the computation basis, the remaining qubits will be encoded in $\mathcal{C}_{BS}^{(n)}$.

Appendix B

Teleporting into Code Blocks

A general method for preparing a code block in the *logical* state $|\psi\rangle$ starting from a single-qubit state $|\psi\rangle$ is to *teleport* into the code block [59]; figure B.1 shows the circuit. First, a *logical* Bell state is prepared by using logical $|+\rangle$ and $|0\rangle$ ancilla blocks and interacting them using a logical CNOT gate. Then, one of the two ancilla blocks is decoded by using the decoding circuit, \mathcal{D} . Finally, the qubit output from the decoder interacts with a single input qubit prepared in the state $|\psi\rangle$ via a CNOT and both qubits are measured as shown in the figure. The *logical* Pauli correction operator, P_L , of the teleportation protocol restores the state of the output code block to be the *logical* state $|\psi\rangle$ as desired.

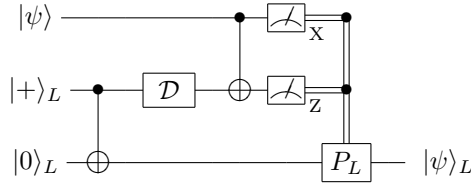


Figure B.1. Teleporting the single-qubit state $|\psi\rangle$ into the code block. First, two ancillary blocks are prepared in a logical Bell state and, then, one of the two blocks is decoded using the decoder, \mathcal{D} . Finally, a measurement along the *Bell basis* $\{(Z^{j_1} X^{j_2} \otimes I)|\Phi_0\rangle \mid j_1, j_2 = 0, 1\}$ is performed on an input qubit prepared in the state $|\psi\rangle$ and the qubit output from the decoder. After the *logical* Pauli correction P is applied, the state of the output code block is $|\psi\rangle_L$ as desired.

We can use this method to prepare the logical $|+i\rangle$ or $|\text{Toffoli}\rangle$ states that are input to the distillation protocols described in §5.3.4. Consider the case where a single-qubit state $|\psi\rangle$ such as $|+i\rangle$ is teleported into the code block at level, say, k of a recursive simulation. If noise is local and stochastic, the probability of a fault during the teleportation of the state into the code block can be upper bounded by

$$p_{\text{anc}} \leq 3p^{(k)} + p_{\text{dec}}^{(k)} + 4p. \quad (\text{B.1})$$

Here $3p^{(k)}$ is an upper bound on the probability of a fault in the level- k logical Bell state (since it is prepared by using three level- k operations), $p_{\text{dec}}^{(k)}$ is the fault probability for decoding one half of the logical Bell state, and $4p$ is an upper bound on the probability of a fault in the preparation of the

input state $|\psi\rangle$ and in the three operations in the measurement along the Bell basis. In general, if a m -qubit state is teleported at level k of a recursive simulation, we can upper bound the probability of a fault with support on a specific one of the m level- k code blocks by replacing the last term, $4p$, in equation (B.1) by $(3 + s)p$ where s is the number of locations in the preparation circuit of the state to be teleported (e.g., for the |Toffoli> state, $m = 3$ and $s = 4$).

We can upper bound $p_{\text{dec}}^{(k)}$ by decoding the level- k block recursively starting from the highest coding level. That is, the decoding circuit, \mathcal{D} , first simulates the code decoder at level k using logical operations at level $k-1$; this results in an output level- $(k-1)$ decoded block. Next, \mathcal{D} simulates the code decoder at level $k-1$ using logical operations at level $k-2$ resulting in a output level- $(k-2)$ decoded block, etc., until the state is decoded to a single qubit. By this procedure and letting D be the number of locations in the decoder,

$$p_{\text{dec}}^{(k)} \leq D \sum_{j=0}^{k-1} p^{(j)}, \quad (\text{B.2})$$

where $p^{(0)} \equiv p$. As long as the strength of stochastic local noise, p , is below the accuracy threshold, $p^{(j)}$ decreases doubly exponentially with j and, thus, our upper bound on $p_{\text{dec}}^{(k)}$ approaches an asymptotic value of order p .

Actually, there is a detail regarding the presence of D in equation (B.2) that we need to discuss. The problem is that, at first, it seems that we need to define a level-1 extended rectangle for the decoding operation that includes a 1-EC followed by a noisy decoder. If this were indeed the case, the probability that this 1-exRec is bad is certainly higher than Dp since we also need to consider the possible occurrence of faults in the leading 1-EC. However, it is easy to see that we do not really need to consider separate extended rectangles for the decoding operation. Instead, by following the intuition about contracting extended rectangles discussed in §4.2.4, we may have the decoder replace the trailing 1-EC of the preceding CNOT 1-exRec, thus creating a 1-conexRec. For example, to be more concrete, this means that in the second logical step in preparing, e.g., a level-1 logical Bell state as in figure B.1, we may consider a *single* 1-conexRec formed by (i) two leading 1-EC steps, (ii) a CNOT 1-Ga, (iii) the decoder acting on the first output block, and (iv) a 1-EC acting on the second output block. We then observe that the probability of badness of this 1-conexRec is upper bounded by $p^{(1)} + Dp$ since if the decoder is executed without faults it is equivalent to a faultless 1-EC followed by a faultless decoder. Indeed, we can obtain this upper bound by considering all fault paths for which the 1-conexRec is bad and the decoder is executed without faults (this probability is at most $p^{(1)}$) *plus* the probability of having even a single fault inside the decoder (this probability is at most Dp). Doing the same contraction of level- k decoders with the preceding CNOT level- k extended rectangles for all $k \geq 2$ implies that the upper bound we obtain by combining equations (B.1) and (B.2) is valid.

Bibliography

- [1] P. Aliferis, D. Gottesman, and J. Preskill. Quantum accuracy threshold for concatenated distance-3 codes. *Quantum Inf. Comp.*, 6(2):97–165, 2006. arXiv e-print quant-ph/0504218.
- [2] P. Aliferis and B. M. Terhal. Fault-tolerant quantum computation for local leakage faults. *Quantum Inf. Comp.*, 7:139–157, 2007. arXiv e-print quant-ph/0511065.
- [3] P. Aliferis and A. W. Cross. Sub-system fault tolerance with the Bacon-Shor code. To appear in *Phys. Rev. Lett.*, 2006. arXiv e-print quant-ph/0610063.
- [4] R. P. Feynman. Simulating physics with computers. *Int. J. Theor. Phys.*, 21:467–488, 1982.
- [5] R. P. Feynman. Quantum mechanical computers. *Optics News*, 11(11), 1985.
- [6] D. Deutsch. Quantum theory, the Church-Turing principle and the universal quantum computer. *Proc. R. Soc. Lond. A*, 400:97–117, 1985.
- [7] D. Deutsch. Quantum computational networks. *Proc. R. Soc. Lond. A*, 425:73–90, 1989.
- [8] Yu. I. Manin. The computable and incomputable. *Soviet Radio, Moscow*, 1980.
- [9] P. Benioff. Quantum mechanical Hamiltonian models of Turing machines. *J. Stat. Phys.*, 29:515–546, 1982.
- [10] A. Peres. Reversible logic and quantum computers. *Phys. Rev. A*, 32:3266–3276, 1985.
- [11] A. Kitaev. Quantum measurements and the abelian stabilizer problem. *Electronic Colloquium on Computational Complexity (ECCC)*, 3:3, 1996. arXiv e-print quant-ph/9511026.
- [12] S. Lloyd. Universal quantum simulators. *Science*, 273:1073–1078, 1996.
- [13] C. Zalka. Simulating quantum systems on a quantum computer. *Proc. R. Soc. Lond. A*, 454, 1998.
- [14] M. H. Freedman, A. Kitaev, and Z. Wang. Simulation of topological field theories by quantum computers. *Commun. Math. Phys.*, 227:587–603, 2002.

- [15] M. Byrnes and Y. Yamamoto. Simulating lattice gauge theories on a quantum computer. 2005. arXiv e-print quant-ph/0510027.
- [16] D. S. Abrams and S. Lloyd. Simulations of many-body fermi systems on a universal quantum computer. *Phys. Rev. Lett.*, 79:2586–2589, 1997.
- [17] D. S. Abrams and S. Lloyd. Quantum algorithm providing exponential speed increase for finding eigenvalues and eigenvectors. *Phys. Rev. Lett.*, 83:5162–5165, 1999.
- [18] A. Aspuru-Guzik, A. D. Dutoi, P. J. Love, and M. Head-Gordon. Simulated quantum computation of molecular energies. *Science*, 309:1704, 2005. arXiv e-print quant-ph/0604193.
- [19] P. Shor. Algorithms for quantum computation: Discrete logarithms and factoring. In *Proc. 35th Annual Symposium on Foundations of Computer Science*, page 124, Los Alamitos, CA: IEEE Computer Society Press, 1994.
- [20] L. Grover. A fast quantum mechanical algorithm for database search. In *Proc. 28th Annual ACM Symposium on the Theory of Computation*, pages 212–219, New York: ACM Press, 1996.
- [21] S. Hallgren. Polynomial-time quantum algorithms for Pell’s equation and the principal ideal problem. In *Proc. of the 34th Annual ACM Symposium on the Theory of Computation*, pages 653–658, New York: ACM Press, 2002.
- [22] M. Sipser. *Introduction to the theory of computation*. Boston, MA: PWS Publishing Company, 1997.
- [23] M. A. Nielsen and I. L. Chuang. *Quantum computation and quantum information*. Cambridge, U.K.: Cambridge University Press, 2000.
- [24] J. Preskill. Lecture notes on Ph219: Quantum computation. Available online at <http://www.theory.caltech.edu/people/preskill/ph229/>.
- [25] A. Yu. Kitaev, A. H. Shen, and M. N. Vyalyi. Classical and quantum computation. American Mathematical Society, 2002.
- [26] B. W. Schumacher. Sending entanglement through noisy quantum channels. *Phys. Rev. A*, 54:2614, 1996. arXiv e-print quant-ph/9604023.
- [27] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. Wootters. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.*, 70(13):1895, 1993.
- [28] E. Bernstein and U. Vazirani. Quantum complexity theory. *SIAM J. Comput.*, 26(5):1411–1473, 1997.

- [29] P. O. Boykin, T. Mor, M. Pulver, V. Roychowdhury, and F. Vatan. On universal and fault-tolerant quantum computing. In *Proc. 40th Annual Symposium on Foundations of Computer Science*, Los Alamitos, CA: IEEE Computer Society Press, 1999. arXiv e-print quant-ph/9906054.
- [30] A. Kitaev. Quantum computations: Algorithms and error correction. *Russ. Math. Surv.*, 52(6):1191–1249, 1997.
- [31] Y. Shi. Both Toffoli and controlled-not need little help to do universal quantum computation. 2002. arXiv e-print quant-ph/0205115.
- [32] D. Gottesman and I. Chuang. Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations. *Nature*, 402:390–393, 1999. arXiv e-print quant-ph/9908010.
- [33] G. Nebe, E. M. Rains, and N. J. A. Sloane. The invariants of the Clifford groups. *Designs, Codes, and Cryptography*, 24:99–122, 2001. arXiv e-print math.co/0001038.
- [34] C. M. Dawson and M. A. Nielsen. The Solovay-Kitaev algorithm. *Quant. Inf. Comp.*, 6(1):81–95, 2006. arXiv e-print quant-ph/0505030.
- [35] W. K. Wootters and W. H. Zurek. A single quantum cannot be cloned. *Nature*, 299:802–803, 1982.
- [36] P. Shor. Scheme for reducing decoherence in quantum computer memory. *Phys. Rev. A*, 52:2493–2496(R), 1995.
- [37] A. Steane. Error correcting codes in quantum theory. *Phys. Rev. Lett.*, 77:793, 1996.
- [38] E. Knill and R. Laflamme. A theory of quantum error correcting codes. *Phys. Rev. A*, 55:900, 1997. arXiv e-print quant-ph/9604034.
- [39] C. H. Bennett, David P. DiVincenzo, John A. Smolin, and William K. Wootters. Mixed state entanglement and quantum error correction. *Phys. Rev. A*, 54:3824–3851, 1996. arXiv e-print quant-ph/9604024.
- [40] A. R. Calderbank, E. M. Rains, P. W. Shor, and N. J. A. Sloane. Quantum error correction and orthogonal geometry. *Phys. Rev. Lett.*, 78:405, 1997. arXiv e-print quant-ph/9605005.
- [41] D. Gottesman. A class of quantum error-correcting codes saturating the quantum Hamming bound. *Phys. Rev. A*, 54:1862–1868, 1996. arXiv e-print quant-ph/9604038.
- [42] D. Gottesman. Stabilizer codes and quantum error correction. PhD thesis, California Institute of Technology, Pasadena, CA, 1997. arXiv e-print quant-ph/9705052.

- [43] A. R. Calderbank and P. W. Shor. Good quantum error-correcting codes exist. *Phys. Rev. A*, 54:1098–1105, 1996. arXiv e-print quant-ph/9605005.
- [44] A. Steane. Multiple particle interference and quantum error correction. *Proc. R. Soc. Lond. A*, 452:2551, 1996. arXiv e-print quant-ph/9601029.
- [45] J. Von Neumann. Probabilistic logics and the synthesis of reliable organisms from unreliable components. In C. E. Shannon and J. McCarthy, editors, *Automata Studies*, pages 43–98, Princeton, NJ: Princeton University Press, 1955.
- [46] P. Gács. Reliable computation with cellular automata. In *Proc. 15th Ann. ACM Symp. on Theory of Computing*, pages 32–41, New York: ACM Press, 1983,
- [47] P. Gács. Reliable cellular automata with self-organization. *J. Stat. Phys.*, 103:45–267, 2001.
- [48] D. Aharonov, M. Ben-Or, R. Impagliazzo, and N. Nisan. Limitations of noisy reversible computation. 1996. arXiv e-print quant-ph/9611028.
- [49] B. M. Terhal and G. Burkard. Fault-tolerant quantum computation for local non-markovian noise. *Phys. Rev. A*, 71(012336), 2005. arXiv e-print quant-ph/0402104.
- [50] D. Aharonov, A. Kitaev, and J. Preskill. Fault-tolerant quantum computation with long-range correlated noise. *Phys. Rev. Lett.*, 96(050504), 2006. arXiv e-print quant-ph/0510231.
- [51] P. Shor. Fault-tolerant quantum computation. In *Proc. 37th Annual Symposium on Foundations of Computer Science*, page 56, Los Alamitos, CA: IEEE Computer Society Press, 1996. arXiv e-print quant-ph/9605011.
- [52] E. Knill, R. Laflamme, and W. Zurek. Resilient quantum computation: Error models and thresholds. *Proc. R. Soc. Lond. A*, 454:365–384, 1998. arXiv e-print quant-ph/9702058.
- [53] C. Crépeau, D. Gottesman, and A. Smith. Secure multi-party quantum computing. In *Proc. 34th Annual ACM Symposium on the Theory of Computation*, 2002. arXiv e-print quant-ph/0206138.
- [54] E. Knill and R. Laflamme. Concatenated quantum codes. 1996. arXiv e-print quant-ph/9608012.
- [55] D. Aharonov and M. Ben-Or. Fault-tolerant quantum computation with constant error. In *Proc. 29th Annual ACM Symposium on the Theory of Computation*, page 176, New York: ACM Press, 1998. arXiv e-print quant-ph/9611025; D. Aharonov and M. Ben-Or (1999), arXiv e-print quant-ph/9906129.

- [56] J. Preskill. Reliable quantum computers. *Proc. R. Soc. Lond. A*, 454:385–410, 1998. arXiv e-print quant-ph/9705031.
- [57] B. W. Reichardt. Fault-tolerance threshold for a distance-three quantum code. 2005. arXiv e-print quant-ph/0509203.
- [58] E. Knill. Fault-tolerant postselected quantum computation: Schemes. 2004. arXiv e-print quant-ph/0402171.
- [59] E. Knill. Quantum computing with realistically noisy devices. *Nature*, 434:39–44, 2005. arXiv e-print quant-ph/0410199.
- [60] E. Knill, R. Laflamme, and G. J. Milburn. Efficient linear optics quantum computation. *Nature*, 409:46, 2001. arXiv e-print quant-ph/0006088.
- [61] T. P. Orlando, J. E. Mooij, L. Tian, C. H. van der Wal, L. S. Levitov, S. Lloyd, and J. J. Mazo. Superconducting persistent-current qubit. *Phys. Rev. B*, 60:15398–15413, 1999.
- [62] D. P. DiVincenzo, D. Bacon, J. Kempe, G. Burkard, and K. B. Whaley. Universal quantum computation with the exchange interaction. *Nature*, 408:339–342, 2000.
- [63] J. R. Petta, A. C. Johnson, J. M. Taylor, E. A. Laird, A. Yacoby, M. D. Lukin, C. M. Marcus, M. P. Hanson, and A. C. Gossard. Coherent manipulation of coupled electron spins in semiconductor quantum dots. *Science*, 309(5744), 2005.
- [64] L. Tian and S. Lloyd. Resonant cancellation of off-resonant effects in a multilevel qubit. *Phys. Rev. A*, 62(050301(R)), 2000.
- [65] M. S. Byrd, D. A. Lidar, L-A Wu, and P. Zanardi. Universal leakage elimination. *Phys. Rev. A*, 71(052301), 2005.
- [66] T. Pellizzari, S. A. Gardiner, J. I. Cirac, and P. Zoller. Decoherence, continuous observation, and quantum computing: A cavity QED model. *Phys. Rev. Lett.*, 75:3788–3791, 1995.
- [67] J. Preskill. Fault-tolerant quantum computation. In H.-K. Lo, S. Popescu, and T. P. Spiller, editors, *Introduction to quantum computation*, pages 213–269. World Scientific, Singapore, 1998. arXiv e-print quant-ph/9712048.
- [68] C. Mochon. Anyon computers with smaller groups. *Phys. Rev. A*, 69(032306), 2004. arXiv e-print quant-ph/0306063.
- [69] R. Raussendorf and H. J. Briegel. A one-way quantum computer. *Phys. Rev. Lett.*, 86:5188, 2001. arXiv e-print quant-ph/0010033.

- [70] M. A. Nielsen. Universal quantum computation using only projective measurement, quantum memory, and preparation of the 0 state. *Phys. Lett. A*, 308:96, 2003. arXiv e-print quant-ph/0108020.
- [71] D. W. Leung. Two-qubit projective measurements are universal for quantum computation. 2001. arXiv e-print quant-ph/quant-ph/0111122.
- [72] D. W. Leung. Quantum computation by measurements. *Int. J. of Quant. Inf.*, 2:33–43, 2004. arXiv e-print quant-ph/0310189.
- [73] F. Verstraete and J.I. Cirac. Valence bond solids for quantum computation. *Phys. Rev. A*, 70(060302), 2004. arXiv e-print quant-ph/0311130.
- [74] P. Aliferis and D. W. Leung. Computation by measurements: A unifying picture. *Phys. Rev. A*, 70(062314), 2004. arXiv e-print quant-ph/0404082.
- [75] A. Childs, D. W. Leung, and M. A. Nielsen. Unified derivations of measurement-based schemes for quantum computation. *Phys. Rev. A*, 71(032318), 2005. arXiv e-print quant-ph/0404132.
- [76] S. Perdrix. State transfer instead of teleportation in measurement-based quantum computation. 2004. arXiv e-print quant-ph/0402204.
- [77] X. Zhou, D. Leung, and I. Chuang. Methodology for quantum logic gate construction. *Phys. Rev. A*, 62(052316), 2000. arXiv e-print quant-ph/0002039.
- [78] M. A. Nielsen. Optical quantum computation using cluster states. *Phys. Rev. Lett.*, 93(040503), 2004. arXiv e-print quant-ph/0402005.
- [79] R. Raussendorf, D. E. Browne, and H. J. Briegel. Measurement-based computation on cluster states. *Phys. Rev. A*, 68(022312), 2003. arXiv e-print quant-ph/0301052.
- [80] C. M. Dawson, H. L. Haselgrove, and M. A. Nielsen. Noise thresholds for optical quantum computers. *Phys. Rev. Lett.*, 96(020501), 2006. arXiv e-print quant-ph/0509060.
- [81] R. Raussendorf. Measurement-based quantum computation with cluster states. PhD thesis, Ludwig-Maximilians Universität München, 2003.
- [82] M. A. Nielsen and C. M. Dawson. Fault-tolerant quantum computation with cluster states. *Phys. Rev. A*, 71(042323), 2005. arXiv e-print quant-ph/0405134.
- [83] P. Aliferis and D. W. Leung. Simple proof of fault tolerance in the graph-state model. *Phys. Rev. A*, 73(032308), 2006. arXiv e-print quant-ph/0503130.
- [84] D. P. DiVincenzo and P. Aliferis. Effective fault-tolerant quantum computation with slow measurements. *Phys. Rev. Lett.*, 98(020501), 2007. arXiv e-print quant-ph/0607047.

- [85] D. Gottesman. Fault-tolerant quantum computation with local gates. *J. Mod. Opt.*, 47:333–345, 2000. arXiv e-print quant-ph/9903099.
- [86] K. M. Svore, D. P. DiVincenzo, and B. M. Terhal. Noise threshold for a fault-tolerant two-dimensional lattice architecture. 2006. arXiv e-print quant-ph/0604090.
- [87] D. P. DiVincenzo and P. Shor. Fault-tolerant error correction with efficient quantum codes. *Phys. Rev. Lett.*, 77:3260, 1996. arXiv e-print quant-ph/9605031.
- [88] A. Steane. Active stabilisation, quantum computation and quantum state synthesis. *Phys. Rev. Lett.*, 78(2252), 1997. arXiv e-print quant-ph/9611027.
- [89] D. Gottesman. Theory of fault-tolerant quantum computation. *Phys. Rev. A*, 57:127, 1998. arXiv e-print quant-ph/9702029.
- [90] D. Gottesman. The Heisenberg representation of quantum computers. In S. P. Corney, R. Delbourgo, and P. D. Jarvis, editors, *Proc. XXII International Colloquium on Group Theoretical Methods in Physics*, pages 32–33, Cambridge, MA: International Press, 1999. arXiv e-print quant-ph/9807006.
- [91] D. Bacon. Operator quantum error correcting subsystems for self-correcting quantum memories. *Phys. Rev. A*, 73:012340, 2006. arXiv e-print quant-ph/0506023.
- [92] B. W. Reichardt and Y. Ouyang. Unpublished.
- [93] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill. Topological quantum memory. *J. Math. Phys.*, 43(4452), 2002.
- [94] D. Gottesman, 2004. Problem 4, Set 5, C&O 639, University of Waterloo. Online at <http://www.perimeterinstitute.com/personal/dgottesman/CO639-2004/>.
- [95] P. Aliferis, D. Gottesman, and J. Preskill. Accuracy threshold for postselected quantum computation. 2007. In preparation.
- [96] D. Poulin. Stabilizer formalism for operator quantum error correction. *Phys. Rev. Lett.*, 95(230504), 2005. arXiv e-print quant-ph/0508131.
- [97] S. Bravyi and A. Kitaev. Universal quantum computation with ideal Clifford gates and noisy ancillas. *Phys. Rev. A*, 71(022316), 2005. arXiv e-print quant-ph/0403025.
- [98] B. W. Reichardt. Error-detection-based quantum fault tolerance against discrete Pauli noise. PhD thesis, University of California, Berkeley, CA, 2006. arXiv e-print quant-ph/0612004.
- [99] S. Bravyi and A. Kitaev. Quantum codes on a lattice with boundary. 1998. arXiv e-print quant-ph/9811052.